DOCUMENT RESUME

ED 194 635 TM 800 767

AUTHOR	Anastasi, Anne: Geisinger, Kurt F.
TITLE	Use of Tests with Schoolchildren. Final Project Report.
INSTITUTION	Fordham Univ., Bronx, N.Y. Dept. of Psychology.
SPONS AGENCY	Harcourt Brace Jovanovich, Inc., New York, N.Y.
PUB DATE	Oct 80
NOTE	181p.
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Achievement Tests: Elementary Education: Elementary School Students: Feedback: Intelligence Tests: Minority Groups: *Parent Attitudes: *Standardized Tests: *Teacher Attitudes: *Testing Problems: *Testing Programs: Test Interpretation
IDENTIFIERS	*Test Directors: Test Publishers: *Test Use

ABSTRACT

How standardized achievement and mental ability tests are used in the schools, and what parents and teachers think about such testing, were investigated. Guided inventories were administered to 207 teachers from 10 schools and to 223 parents from 12 Parent-Teacher Association groups. Intensive individual interviews were conducted with 15 testing coordinators with responsibility for district-wide testing decisions. Teacher inventory covered personal data, testing activities and use of test scores, test opinions, and familiarity with testing concepts and interpretive context. Parent inventory covered personal data, attitudes toward testing in general and toward specific educational uses of tests, knowledge about uses of tests in own child's school, reactions to own child being tested, and extent and nature of feedback received about child's test performance. Results for teachers, testing coordinators, and parents are reported in separate chapters. A discussion section integrates findings from these parts and presents implications, with special focus on testing communications at several levels. Teacher inventories, parent attitude measures, and test coordinators' interview questions are appended. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED194635

USE OF TESTS WITH SCHOOLCHILDREN

ANNE ANASTASI and KURT F. GEISINGER

Fordham University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Geisinger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800767

OCTOBER 1980

USE OF TESTS WITH SCHOOLCHILDREN

Final Project Report

Anne Anastasi and Kurt F. Geisinger
Department of Psychology
Fordham University

October 1980

PROJECT STAFF

Principal Investigator	Anne Anastasi
Project Director	Kurt F. Geisinger
Faculty Research Associate	Lawrence A. Bradley

Graduate Assistants:

Vincent Hevern, S.J.
Cynthia Kabat
Gardner Locke
Jean Powell-Kirnan

Secretary: Peggy Mackiewicz

CONSULTANTS

Anthony N. Baratta
Eileen Byrne
John C. Houtz
Joseph F. X. McCarthy
Vincent M. Novak, S.J.
Lillian R. Restaino
Jay S. Sexter
John M. Skalski
Thomas G. Vinci
John F. Walsh
Steven Wepner

This project was supported in part by a grant from Harcourt Brace Jovanovich, Inc., to Fordham University.

The opinions expressed in this report are those of the authors and do not reflect the views of Fordham University or Harcourt Brace Jovanovich Inc.

We want to express our deep appreciation to the school principals, testing coordinators, presidents of Parent-Teacher Associations, and the many individual teachers and parents, whose interest and cooperation made this project possible.

ABSTRACT

The principal object was to investigate how standardized achievement and mental ability tests are used in the schools, and what parents and teachers think about such testing. Guided inventories were administered to 207 teachers from 10 schools and to 223 parents from 12 PTA groups. Intensive individual interviews were conducted with 15 testing coordinators with responsibility for district-wide testing decisions. Teacher inventory covered personal data, testing activities and use of test scores, test opinions, and familiarity with testing concepts and interpretive context. Parent inventory covered personal data, attitudes toward testing in general and toward specific educational uses of tests, knowledge about uses of tests in own child's school, reactions to own child being tested, and extent and nature of feedback received about child's test performance. Results for teachers, testing coordinators, and parents are reported in Parts I, II, and III, respectively. A discussion section integrates findings from these parts and presents implications, with special focus on testing communications at several levels.

TABLE OF CONTENTS

Background of the Study..... 1

The Present Study..... 3
Objectives..... 3
General Plan..... 5

Part I: Teachers..... 8
Procedure..... 8
Participant characteristics..... 12
Total group results..... 17
Comparative analysis of responses of black teachers.... 37

Part II: Testing Coordinators..... 49
Procedure..... 49
Participant characteristics..... 51
Interview findings..... 54

Part III: Parents..... 76
Procedure..... 76
Participant characteristics..... 79
Total group results..... 81
Comparative analysis of responses of black parents.....104
Comparative analysis of parent and teacher responses on Test Opinion Inventory.....110

Discussion.....113
Overview.....113
Communication in relation to standardized testing and test results.....115
The special needs of teachers.....116
The concerns of parents.....121
The school in a societal context.....125
The role of test publishers.....128

Appendix.....131
A. Sample letters regarding confidentiality commitment.133
B. School Personnel Inventory.....137
C. Interview Schedule for Testing Coordinators.....149
D. Parent Survey.....155

References.....167

BACKGROUND OF THE STUDY

There are several indications that testing practice may undergo significant changes in the 1980s (Anastasi, 1976; Ch. 12; 1979, Ch. 2; 1980; Feuerstein, 1979; Lerner, 1979, 1980; Maloney & Ward, 1976; Messick, 1980; Staats, 1970). Some of these changes concern the purposes for which tests are used, such as selection vs. placement, prediction vs. diagnosis, and institutional vs. individual decisions. Some will involve major revisions in the nature of tests themselves or the development of new tests following innovative approaches. It appears likely, however, that among the most significant changes will be those pertaining to the interpretation of test results. With the dramatic expansion in test development and use since the 1940s, test constructors as well as test users have tended to focus more and more on testing technology and to become dissociated from the mainstream of behavioral science (Anastasi, 1967).

Advances in psychology, genetics, and other relevant fields have not been adequately reflected in the way test scores are interpreted and used in the decisions that affect individuals. Outworn concepts and models of the development of human behavior have survived tenaciously in the interpretation of test scores (Cravens, 1978). In the 1970s, several test authors and publishers made feeble attempts to dislodge some of the excess

meanings that had become associated with their tests by replacing the term "intelligence" in the test names with more neutral terms. But the public -- and all too many professionals-- still retain an "IQ" orientation and treat a mental test score as a property of the organism.

Current criticisms of testing by various public interest associations, minority-group spokespersons, and other organized groups stem from a diversity of sources. Political motives obviously play a large part (Lerner, 1979, 1980). Public media are attracted by the news value of the ensuing controversies and may create a distorted impression of the prevalence of negative attitudes toward testing. To some extent, however, the criticisms are directed against actual misuses and misinterpretations of tests; and insofar as this is true, the objectionable practices should be identified and corrected. It is the responsibility of both test users (such as school systems) and test publishers to take steps to correct these misuses.

In the 1970s, individual investigators as well as professional and scientific organizations began to examine intensively the ethical, societal, and scientific aspects of test uses and misuses. Their approaches vary widely, ranging from public opinion polling (Brim, 1965; Brim, Glass, Neulinger, & Firestone,

1969; Goslin, 1967) to historical, sociological, and ethnographic studies (Resnick, 1979; Resnick & Resnick, 1978; Airasian, 1979; O'Regan, Airasian,& Madaus, 1979). Psychologists from diverse specialties, ranging from experimental and psychometric to social and clinical, have amassed extensive bodies of data on such problems as test anxiety (Spielberger et al., 1978; Tryon, 1980), the effects of tests on students (Kirkland, 1971), test fairness and bias (Breland, 1979; Gross & Su, 1975; Hunter & Schmidt, 1976; Hunter, Schmidt, & Hunter, 1979; Jensen, 1980, Chs. 9 & 10; Linn, 1978; Petersen & Novick, 1976), the effects of special training programs on test scores (Babad & Budoff, 1974; Budoff & Corman, 1974; College Board, 1979; Feuerstein, 1979), the role of cultural differences in test performance (Berry, 1972; Cole & Bruner, 1971; Goodnow, 1976; Neisser, 1976, 1979), and the nature of the constructs assessed by tests (Construct Validity, 1980; Messick, 1980).

THE PRESENT STUDY

Objectives

The major purpose of this project was to gather data on how tests are actually used with schoolchildren, and what the people who are closely involved with childrem know and think about tests. Prominent among such people, of course, are teachers (and other school personnel) and parents. An important

component of current dissatisfactions with testing pertains to test use with ethnic and cultural minorities. Accordingly, we were especially concerned with how tests are used with minority children, and with the views of minority parents and educators regarding test use. At the same time, our approach was deliberately broad. Tests can be effectively used to benefit all children or they can be misused to the detriment of all children. The use of tests with minorities can be best understood -- and improved -- within the context of test use with all children.

A major hypothesis that underlies the present study is that the communication of test results to teachers and parents leaves much to be desired. Certain parts of the study are concerned with this highly specific type of information transmittal. Another relevant aspect of communication pertains to the dissemination of general knowledge about testing and the meaning of test scores. Today's test consumers (e.g., teachers, parents, test-takers) want comprehensible explanations about testing so that they can judge its value for themselves. And this is a goal that <u>can</u> be achieved, as illustrated by two recent examples of effective printed communications, one directed to the parents of schoolchildren (Dyer, 1980) and one to college-bound high school seniors (College Board, 1980). There is also need for

communication in the opposite direction, namely from parents, teachers, and other test users to educational administrators and test authors and publishers. Such communication can reveal current misuses and misinterpretations that are undermining the appropriate and constructive use of tests.

General Plan

The present study followed a three-pronged approach to explore test use in the schools, with special concentration on the elementary school level. Data were collected from three sets of persons critically involved with the use of educational tests: (a) teachers, (b) testing coordinators with major responsibility for school testing programs, and (c) parents of schoolchildren. Data from teachers and parents were gathered through written questionnaires administered by project staff to small groups, with opportunity for discussion at the end of each session. Information from testing coordinators was obtained through intensive individual interviews conducted by a senior member of the project staff.

The major questions under investigation may themselves be grouped into three categories, although data bearing on any one category were usually derived from more than one type of participant. First, we were interested in how tests are selected and how decisions about testing programs are made. Second, we wished to explore the purposes for which tests are used in the

schools and the ways in which they contribute to educational decisions about individual children. Third, we wanted to identify strengths and weaknesses of current test usage as perceived by parents and school personnel.

At the outset, it was necessary to address fully and explicitly the problem of protecting the confidentiality of project data. Not only is this a general ethical obligation in conducting all scientific research, but it also soon became apparent that formal assurance regarding the strict confidentiality of group identity was a prerequisite for participation in our study. Two examples of the type of written commitment required from us are included in Appendix A. Following the wording prescribed by the participating official, in one of these letters we agreed "that in no way will any individuals in your school, your school itself, the school district, town, county, etc. ever be identified in any way. You will have complete anonymity." In the other appended letter, we promised "that no mention of any of the individuals in the study, of the school in question, or their location, proximate or general, will be made." It might be added that, because of the rather extreme precautions taken to protect the identity of individuals and institutions, we were able to obtain sensitive data that might otherwise have been unavailable to us. We also feel free to report our findings

fully, without glossing over what could have been embarassing details. The multiplicity of our data sources served as a means of safeguarding the identity, not only of persons and schools, but also of the various publishers whose tests were cited by the respondents.

Most of our contacts, especially for teacher and parent groups, were made through the cooperation of our project consultants at the Fordham University School of Education, whose students are drawn from a large and diversified set of schools. Other contacts with participating individuals and schools were made through professional persons in various school systems known personally to members of the project staff. Within the limitations imposed by the confidentiality requirement, we can report that our data were obtained in a tristate area within commuting distance of New York City.[1] The teachers, parents, and testing coordinators were drawn from settings representing a wide range in type of school (public, private, or parochial, general or special-purpose), type of neighborhood (inner-city, metropolitan, suburban), income level, and ethnic composition.

In this report, the methodology and results for those parts of the study concerned with teachers, testing coordinators, and parents are described in separate sections, in that order. A

---

[1] With one exception, a school in a large northeastern city outside this area.

final, discussion section draws from all three parts of the study.

PART I: TEACHERS

Procedure

Development and Pretesting of School Personnel Inventory. An instrument for small-group administration to teachers and other appropriate school personnel was developed, pretested, and revised during the first project year (1978-79). Pretest data were collected from 10 classes at the Fordham University Graduate School of Education, nine held at the Lincoln Center campus and one at the Tarrytown campus. All but one of the classes were in the Curriculum Division, the tenth was in the Administration Division. Completed forms were obtained from a total of 124 persons. The respondents appeared to be typical of graduate students in a College of Education; the majority were currently employed as teachers, and a sizeable proportion were educational administrators.[2] Statistical analyses of the responses, together with a consideration of comments by respondents and recommendations by project consultants, led to the preparation of the final form of the School Personnel Inventory,

[2] Supplementary pretest data on one portion of the instrument, the Test Opinion Inventory, were also obtained from 104 students in a Liberal Arts college.

reproduced in Appendix B.

The entire inventory is divided into four parts: a one-page questionnaire on background information, the Test Activities Inventory, the Test Opinion Inventory, and the Test Usage Inventory. The Test Activities Inventory consists of two parallel sets of questions covering the same information, but with regard to achievement tests in the one and mental ability tests in the other. In both cases, the questionnaire begins with a brief characterization of the type of test under consideration (achievement or mental ability), followed by a list of 15 possible activities that a teacher could perform in relation to these tests. The respondents circle whether they perform each of the activities regularly, occasionally, or never. A final category covers any other activity performed, to be filled in by the respondent and marked for frequency in the same way. Each of the two sections of the Test Activities Inventory concludes with a single nine-point rating scale, on which the respondent evaluates how useful he or she has found that type of test, and two open-ended questions. The first question asks what could be done to make the tests more useful; the second requests the respondent to name one or more specific examples of achievement or mental ability tests, respectively. The latter question was included largely as a check on what these two common test categories meant to the respondent.

The Test Opinion Inventory consists of 15 nine-point Likert-type attitude items. Each item presents a statement that is either clearly favorable or clearly unfavorable to current educational uses of tests. These statements were drawn principally from opinions about tests expressed in the media or in other popular discussion of tests by individual spokespersons or organized groups. We were interested chiefly in obtaining reactions to adverse criticisms about test use, and accordingly nine of the 15 statements were of this type. In order to minimize the operation of an acquiescence response set (or its reverse), and to ensure that each statement was carefully read and evaluated independently, we inserted six statements favorable to testing. These statements were placed first and last in the list and in four intermediate, randomly distributed positions.

The Test Usage Inventory was designed to assess teachers' understanding of certain basic testing concepts, as well as their familiarity with relevant behavioral knowledge that would affect the interpretation and use of test scores. In its pretesting stage, this inventory was analyzed in terms of internal consistency (coefficient alpha), item difficulty index, item-test correlation (corrected for item effect on total score), and proportion of persons selecting each option. On the basis of the pretest data, several items were deleted, the options of other items were revised, and new items were added. The final version consists of 15 five-

option multiple-choice items, with a "Don't Know" option in fifth position.

Data-Gathering Procedures. The School Personnel Inventory was administered by a member of the project staff to teachers from 10 schools. All were elementary schools, except for one junior high school from which we obtained only nine teachers. The schools included eight public schools and two religious-affiliated private schools. A total of 207 completed inventories were obtained, the number from each school ranging from 8 to 38.[3]

The typical sequence of steps followed in gathering data from a school began with a telephone call to the school principal, whose name had been obtained through one of our referral sources. This call usually led to a meeting with the principal in his or her office, during which a project staff member explained the nature and purpose of the project, outlined the benefits the school could derive from participation, and discussed the arrangements for protecting confidentiality. In some cases, at the principal's request, a similar meeting with the district superintendent followed. Next came more phone calls to settle on a data for actual data gathering at the school, and in some cases the submission of letters formalizing the confidentiality arrangements. The School Personnel Inventory was administered

[3]Since two of the schools had a single principal and regularly held joint faculty meetings, 59 teachers from the two schools were tested in a single session.

during a regularly scheduled faculty meeting at each school. Typically, the project staff member addressed the faculty for 10 to 15 minutes regarding the nature and purpose of the project, and then administered the inventory. The session ended with an open discussion of any questions the participants wished to raise.

After the completion of data analysis, in October 1980, each school received an individually computerized report of its own data, together with the overall project results on the School Personnel Inventory and an invitation to contact the project staff if further discussion of the report was desired.

Participant Characteristics

The one-page questionnaire included in the School Personnel Inventory provided the data for the description of participant characteristics summarized in this section. These data cover: (1) demographic characteristics of the teachers themselves, and (2) information regarding the extent and nature of student contacts the teachers had in their normal school activities.

Demographic Data. Table 1 summarizes the personal characteristics of the 207 teachers who participated in this portion of the study. The sample was predominantly female (81%). The largest proportion were in their 20s and 30s, but about 40% were distributed through the 40s and 50s, and 3% were 60 or over. Ethnically, the large majority (84%) were white (non-Hispanic), with 13% black and only 1.5% Hispanic. Educationally, the largest group (38%) had a Master's degree plus some

Table 1
Teacher Sample: Demographic Data
(N=207)[a]

Variable	Percentages			
Sex	Female	80.8	Male	19.2
Age	20-29	27.0	50-59	19.5
	30-39	29.0	60 & up	3.0
	40-49	21.5		
Ethnic	White(non-Hispanic)	84.1	Hispanic	1.5
	Black	13.4	Amer.Indian	0
	Oriental	0	Other[b]	1.0
Education	H.S.& further	3.4	Master degree	18.8
	Bachelor degree	11.6	Master & further	37.7
	Bachelor & Graduate	28.5	Doctoral degree	0
Testing Training	Graduate course	42.9	Grad & In-service	4.4
	Undergrad.course	20.7	Undergrad & In-service	4.4
	In-service	3.9	Grad.& Undergr. & In-serv.	6.4
	Grad. & Undergrad.	6.9	No formal training	10.3
Reported job title	Classroom teacher	67.2	Teacher/Administrator	2.5
	Special educ.teacher	20.2	Administrator	2.5
	Specialty teacher (e.g.,reading,math)	4.5	Social worker[c]	2.0
			Nurse[c]	1.0

[a] For item responses, N=200-207

[b] 1 East Indian, 1 Filipino

[c] Served on Child Study Teams assessing individual students on basis of test scores and other data.

further graduate work; another 19% reported only a Master's degree, while none had a doctoral degree. A bachelor's degree plus some graduate work accounted for another 29%. A relatively small proportion (12%) had only a bachelor's degree, and a few (3%) reported still less education. On the whole, the educational level of this group of teachers was high and fairly typical of that found in large school systems in urban and suburban areas.

Of more interest for the present study is the extent of training in educational and psychological testing. The largest proportion (43%) had had only a graduate course in this area, and the second largest (21%) reported only an undergraduate course. Few (4%) reported only in-service training on testing. Various combinations of two or all three types of training were reported by 22%. Only 10% had received no formal training in educational or psychological testing. With regard to current job title, the sample was composed predominantly of classroom teachers (67%). An additional 20% were special education teachers, and 4½% were teachers of subject-matter specialties, such as reading or mathematics; 5% were functioning as administrators or teacher/administrators. The sample also contains four social workers and two school nurses. These individuals were included because they served on Child Study Teams and hence had considerable experience in the utilization of test

results in assessing individual students and making recommendations about educational placement and adaptive instructional programs.

Student Contacts. Nearly the entire sample (98%) reported having direct contact with students in their daily work. The grade level ranged from pre-kindergarten through high school,[4] but the group taught predominantly at the elementary grades, with the largest number clustering at grades 1 through 6. The number of students with whom our participants had direct contact ranged widely. The extremes of zero and several hundred were represented by administrators, specialty teachers, and other persons with special assignments. The distribution is highly skewed, with clustering well below the center of the range and a median of approximately 34 students.

With regard to ethnic distribution of students, the large majority of participants had contact with predominantly white (non-Hispanic) groups of students. The median number of students in this category was 23.5, while the median number of black students was slightly over 2. The number of black students reported by individual respondents, however, varied widely, ranging from zero to 200. The median numbers of students in other ethnic groups were all under 1, but the individual frequencies also revealed wide diversity among respondents. For

[4]Although we did not gather data in any senior high school, some of our participants included high school students, among others, in their reported student contacts, because of special assignments or teaching at more than one grade level.

Table 2
Student Contacts

Variable	Response Summary			
Direct Contact Grade Levels	Yes 97.6%		No 2.4%	
	Pre-Kindergarten to 12			
Number of students[a]	Mdn = 33.5		Range 0-750	
Mdn N in Ethnic Groups[b]	White	23.5	Black	2.43
	Hispanic	0.30	Amer.Indian	0.003
	Oriental	0.19	Other	0.02

[a] 200 responses

[b] 190 responses

example, 118 teachers reported no contact with Hispanic students, but 18 teachers reported contact with 6 to 33 Hispanics. Similarly, 138 teachers reported no contact with Orientals, but 14 teachers reported contact with 6 to 20 students in this ethnic category. In summary, while the majority of our participants had school contacts with predominantly white (non-Hispanic) students, there was a substantial number of individual participants who had experience with ethnic minorities in their normal school activities.

Total Group Results

Teachers' Perceptions of Achievement and Mental Ability Tests. In soliciting teacher responses regarding both uses and evaluations of standardized tests, we have separated achievement from mental ability tests throughout the inventory. We reached this decision because of the prevalence of this distinction in educational usage, regardless of its theoretical weaknesses. Several of our findings corroborated the need for such a differentiation when seeking information from teachers. At this point, let us examine the teachers' responses to the open-ended question calling for one or more examples of standardized achievement tests with which the respondent had recently been involved, and the corresponding question regarding mental ability tests. The tests named by the respondents provide a clue to their per-

ception of tests belonging in these two categories.

The general results indicate a clear distinction that follows the traditional pattern. Among the achievement tests listed were such well-known batteries as the Iowa Tests of Basic Skills, SRA Achievement Series, Stanford, Metropolitan, and California Achievement Tests, in order of frequency. Diagnostic tests were mentioned especially often, the most common examples including Woodcock-Johnson KeyMath Diagnostic Arithmetic Test, Peabody Individual Achievement Test, and Wide Range Achievement Test, as well as the diagnostic parts of the previously mentioned batteries. A sizeable group of respondents simply mentioned "criterion-referenced tests," without specifying particular instruments. There were a few individual confusions, such as one respondent naming the Stanford-Binet and one the WISC, but these instances were rare. Nearly all respondents listed tests traditionally classified in the achievement category.

Among examples of mental ability tests, the well-known intelligence tests clearly topped the list, with WISC-R, WPPSI, and Stanford-Binet accounting for 62 responses. Among group tests, the Cognitive Abilities Test was cited most often. Others named by three or more respondents included some of the SRA series, the Detroit Tests of Learning Aptitude, and the Otis-Lennon. The concern with learning disabilities is reflected in the inclusion of the Illinois Test of Psycholinguistic Abilities, the Slosson, and

the Bender Gestalt. There is some indication that mental ability tests tended to be identified with individually administered, clinical instruments. For instances, several persons listed the PIAT and the WRAT as mental ability tests -- of course, the PIAT may have been confused with the Peabody Picture Vocabulary Test, which others also listed. Again, however, we can conclude that, apart from a few individual confusions, the large majority of our respondents shared the traditional perception of a mental ability test.

Teachers' Roles with Regard to Tests. Table 3 summarizes the involvement of teachers in testing functions, as well as the way teachers use test scores. For both achievement and mental ability tests, the table shows the percentage of teachers reporting that they engaged regularly, occasionally, or never in each test-related activity. It is evident that this involvement was much greater with achievement than with mental ability tests. The percentage reporting that they regularly performed each listed activity is consistently greater for achievement than for mental ability tests. This difference applies not only to such functions as the selection or administration of tests, but also to the uses of test scores for each of the specific purposes listed in Table 3.

There is nevertheless considerable correspondence across both types of tests in the relative frequency with which each

Table 3
Percentage of Teachers Reporting Each Test Activity

Activity	Achievement Tests[a]			Mental Ability Tests[b]		
	Regularly	Occasionally	Never	Regularly	Occasionally	Never
Authorize testing program	10.4	16.4	73.1	6.8	7.8	85.4
Select tests	14.4	19.4	66.2	5.7	5.7	88.6
Coordinate, manage program	19.7	15.3	65.0	9.9	9.4	80.7
Decide to continue program	10.9	12.4	76.7	6.7	7.2	86.1
Orient students to tests	44.1	30.7	25.2	20.2	18.7	61.1
Administer tests	59.3	28.9	11.8	22.8	20.7	56.5
Receive test scores	72.1	17.4	10.4	49.0	24.2	26.8
Have access to scores	80.3	11.8	7.9	58.0	23.8	18.1
Use scores: understand students	58.3	34.3	7.4	35.1	41.2	23.7
Use scores: instruction	53.4	36.3	10.3	29.5	42.0	28.5
Use scores: grouping students	41.4	43.3	15.3	23.2	40.7	36.1
Use scores: evaluate performance	48.5	37.1	14.4	23.8	40.9	35.2
Use scores: identify special needs	55.1	34.6	10.2	32.6	39.9	27.5
Counsel students re score	20.5	34.5	45.0	8.8	22.3	68.9
Explain scores: parents or others	37.1	43.6	19.3	17.3	32.5	50.3
Other uses	12.4	24.2	63.4	8.0	10.2	81.8

[a] N=201 - 205, but "other uses" N=153

[b] N=191 - 193, but "other uses" N=137

activity is regularly performed. The most frequent teacher roles reported for both types of tests are: (1) having access to scores, and (2) receiving scores routinely. Administering tests ranks third for achievement tests but drops to eighth place for mental ability tests. This discrepancy is understandable since, at least in some of the public schools within our sample, group intelligence tests are not used. Hence the mental ability tests would be individual tests that teachers are not ordinarily trained to administer.

Among the uses of test scores, the relative frequency reported for the two types of tests shows close correspondence. The most common uses for both, in order of frequency, are as an aid in: understanding students, planning and adapting instruction, grouping students for instructional purposes, evaluating individual or group performance, and identifying children with special needs. Also among the ten most common teacher activities for both types of tests are orienting students to taking tests and explaining scores to parents or school personnel. Counseling students about scores was not common, but was reported by relatively more teachers for achievement than for mental ability tests. Relatively few teachers reported decision-making responsibilities regarding testing programs or specific tests. Similarly, few were involved in the coordination or management

of a testing program.

If we compare the percentage of teachers reporting each test activity regularly or occasionally, we find that, for achievement tests, more teachers engage in these activities on a regular than on an occasional basis, especially for the high frequency functions. A clear reversal of this tendency, however, is found in the uses of mental ability scores. For all the uses (last eight items in Table 3), the percentage is consistently larger for occasional than for regular. These differences are understandable insofar as mental ability tests (and especially individual intelligence tests) are likely to be given only in special cases and for special purposes.

It will be noted that the last item in Table 3 refers to "other uses," not listed in the table. This item was marked by 153 respondents for achievement tests and by 137 for mental ability tests. With rare exceptions, the uses specified could be subsumed under one or another of the tabled categories and often represented specific illustrations of a listed use. It is also noteworthy that, with very minor exceptions, the same supplementary uses were submitted for achievement and mental ability tests, although the frequencies were again higher for achievement tests. The most common uses were mentioned about twice as often for achievement as for ability tests. Nevertheless,

the relative frequencies of the corresponding uses were in the same rank order for both types of tests. The most common was the use of tests in individualizing instruction and in writing Individualized Educational Programs (IEPs). Next in order was placement, especially for the handicapped and the accelerated. This was followed by planning instruction and treatment, and by identifying the learning disabled, mentally retarded, and gifted students. It can be seen that all these uses are closely related to each other and to some of the categories listed in Table 3. Explaining scores to parent, or child, or both was listed by three respondents for both types of tests. Two uses, each listed by a single respondent for achievement tests only were "to support grading" and "to determine awards." It is apparent that the write-in responses on test activities served chiefly to provide specific examples and corroboration of the results summarized in Table 3.

Perceived Usefulness of Standardized Tests. Table 4 summarizes the findings from the first of the nine-point rating scales used in several parts of the study. For each scale, respondents are instructed to circle the one number that corresponds to their judgement. In this case, they are asked to indicate how useful they have found standardized achievement tests, in one scale, and ability tests, in the other. The scales range from "not at all useful" (1) to "highly useful" (9).

Table 4
Teacher Ratings of Usefulness of Standardized Tests

Type of Test	Mean	SD	Range	σ_M
Achievement	5.11	2.03	1-9	.14
Mental Ability	4.47	2.16	1-9	.16

Reference to Table 4 shows that opinions varied widely for both types of tests. The full range of the scale was used (1-9), and the standard deviations of the ratings are approximately the same for achievement and ability tests. Mean ratings are close to the center of the scale (5), but opinions tend to be slightly more favorable for achievement than for ability tests, the difference averaging about half a point. While not large, this difference is significant at $p < .01$.

Following each rating for judged usefulness, respondents were asked "What could be done to make these tests more useful to you?" In response, 92 persons submitted one or more suggestions for achievement tests and 22 did so for ability tests. Again there were several common responses for both types of tests, although with lower frequencies for ability tests. It is of particular interest that recommendations regarding the improvement of feedback had the highest frequencies for both types of tests. Specific suggestions in this category included further breakdown of scores; the use of simpler, clearer language for explanations to teachers, students, and parents; and receiving the test results faster and earlier. Another, related set of comments, called for more explanation of the tests themselves, their meaning and coverage. Still other comments submitted by several respondents for both achievement

and ability tests concerned the need for updating and for improving the validity and appropriateness of tests for their uses. Several recommended the use of tests in conjunction with other sources of data about the individual. There were requests for more flexibility in the use of test results, greater access to scores, more psychologists and other trained personnel, and proper testing conditions.

Some recommendations were specific to achievement tests. The largest number referred to the need for tests geared to special populations with regard to their procedures and materials. Other recommendations were for tests more appropriate for the school curriculum, and for more teacher input into test content. Some respondents wanted less culturally biased tests, and some wanted more criterion-referenced tests. Along the same lines, more prescriptive feedback to aid instruction was recommended. In reference to ability tests, one respondent suggested the teaching of test-taking strategies.

Test Opinion Inventory. The items in the Test Opinion Inventory were drawn from a collection of adverse statements about tests culled principally from the media and from the literature distributed by antitest activist groups. The final set includes 15 statements, among which were inserted six statements favorable to tests, in order to encourage careful reading and minimize the operation of a generalized response set. Each

Table 5
Teacher Responses to Test Opinion Inventory

Item No.	Key Topic	Mean[a]	SD
* 1	Ability tests help instruction.	4.43	2.20
2	Testing and human dignity.	3.22	2.13
3	Interviews vs. tests in college admission.	6.26	2.06
4	Testing too widespread in our society.	6.28	2.13
* 5	IQ tests and instructional decisions.	4.48	2.15
6	Tests underpredict minority school achievement.	6.28	2.20
* 7	Reading tests aid teachers.	6.13	1.95
8	Tests measure rote memory.	4.62	2.11
* 9	Test scores independent of appearance.	4.83	2.32
* 10	Tests, cultural handicaps, and remediation.	4.30	2.42
11	Mental ability tests unfair to minorities.	5.59	2.52
12	Eliminate all standardized tests in education.	3.62	2.35
13	National norms encourage competition.	5.05	2.52
14	Measuring only a few traits is harmful.	5.33	2.42
* 15	Proper test use prevents unfair discrimination.	6.37	2.00

Note. Scale 1-9, obtained range 1-9 for each item. Low scores indicate favorable attitude, except on starred items.
For item responses, N=198-204.

[a] σ_M=.14 - .18

statement is rated on a nine-point scale ranging from "Strongly Disagree" (1) through "?" (5) to "Strongly Agree" (9). Table 5 shows the means and standard deviations of the ratings given to each statement. Low scores indicate a favorable attitude toward tests (i.e., rejection of an adverse statement), except for the six starred items, in which high scores indicate a favorable attitude. A middle or neutral position corresponds to a rating of 5. Since the standard errors of the obtained means range from .14 to .18, a conservative estimate of a statistically significant deviation ($p < .01$) from a theoretical value of 5 is indicated by a difference of approximately $\pm$ 0.5 or more. Accordingly, we shall examine items whose means fall outside the 4.5 to 5.5 range.

First, we note the wide range of individual differences in the responses. For each statement, the ratings assigned by individual respondents cover the full scale range from 1 to 9. The SDs vary from 1.95 to 2.52. Turning to the means, we find four that do not differ significantly from the neutral value of 5 (Items 8, 9, 13, 14). Four show <u>favorable</u> reactions to tests. Each of these yields a substantial deviation from 5, their means falling below 4 or above 6. Two represent strong rejections of extreme antitest views, one labeling testing as an insult to human dignity (Item 2), the other advocating the elimination of all standardized testing in the educational system

(Item 12). The other two show strong acceptance of positive statements about tests, one referring to the usefulness of standardized reading tests (Item 7), the other to the value of properly used tests in preventing unfair discrimination (Item 15).

Seven means suggest unfavorable reactions to tests. Of these, three represent weak rejections of positive statements (Items 1, 5, 10), with means slightly below 4.5 (4.30 - 4.48). One represents a weak acceptance of an adverse statement (Item 11), with a mean barely above 5.5 (5.59). The remaining three mean ratings indicate strong acceptance of adverse statements, all falling above 6.25. These three items state that: in the admission of college students, interviews should receive more weight than tests (Item 3); testing is too widespread in our society (Item 4); and tests tend to underpredict the school achievement of minority children (Item 6). It is noteworthy that these statements are not among the most extreme statements in the inventory, nor the most extreme views encountered in the media.

In general, it appears that this highly heterogeneous group of teachers responded in a thoughtful, careful manner. While endorsing several adverse criticisms of tests, the majority view was moderate. Moreover, the response pattern was balanced and meaningful, with rejection of the most extreme assertions

and endorsement of the more moderate. There is also some indication that the respondents differentiated between proper and improper uses of tests, and were more strongly opposed to potential or assumed misuses.

Familiarity with Test Concepts and Interpretive Background. The 15-item, multiple-choice test labeled "Test Usage Inventory" was designed to explore teachers' knowledge about the meaning and interpretation of test scores. Each item is in the form of a realistic question about the implications of a test score, which could arise in a school setting. Four response options are given for each item, plus a fifth "Don't Know" option. Apart from sampling knowledge about types of scores, norms, reliability, and other elementary psychometric concepts, we were interested in assessing the prevalence of certain popular misconceptions regarding the interpretation of test scores and the functions of tests.

As seen in Table 6, the mean number of items correct was 7.74, or slightly over 50%. Inspection of the percentage of persons choosing each response option shows that every wrong option was chosen by someone, with a single exception: Option d, Item 10, was rejected by all respondents. This response option claimed that an intellectual difficulty attributed to past environmental handicaps will be outgrown after the child

Table 6
Teacher Knowledge about Test Concepts and Interpretations

Item No.	Central Idea	Percentage Choosing Each Response Option				
		a	b	c	d	e[a]
1	Error of measurement and score bands.	7.0	18.6	**58.8**	2.5	13.1
2	Need for norms in interpreting numerical scores.	4.5	30.2	5.9	**50.0**	9.4
3	Nature of criterion-referenced tests.	34.5	4.9	1.0	**42.4**	17.2
4	Meaning of grade-equivalent scores.	1.5	3.0	**66.5**	19.2	9.9
5	Meaning of age norms.	10.0	**66.0**	4.0	7.5	12.5
6	Need for supplementary information.	1.0	6.5	**85.6**	1.5	5.5
7	Nature of reading readiness tests.	17.0	4.0	**59.5**	3.5	16.0
8	Probability of individual exceeding median of overlapping distribution.	6.3	**36.1**	3.1	5.2	49.2
9	How norms are established.	**64.7**	14.4	1.5	3.0	16.4
10	Environmental handicaps and test scores.	2.0	**41.3**	36.8	0	19.9
11	Meaning of standard scores.	11.7	39.1	14.2	**17.3**	17.7
12	What mental ability tests measure.	**23.4**	11.9	50.7	3.5	10.5
13	Meaning of percentile scores.	19.1	**43.7**	2.0	16.6	18.6
14	Meaning of test reliability.	**57.5**	3.5	21.5	5.0	12.5
15	Interpretation of grade-equivalent score.	1.0	1.0	1.5	**88.0**	8.5
	Mean Number of Items Correct = 7.74					

Note. Correct responses are boxed
Response Option e = Don't know

is placed in a normal school environment. We can be thankful for this bit of undisputed insight throughout our sample!

Turning now to the correct options, we find that the largest percentage of respondents chose the correct option for all items except Items 11 and 12. The first reveals only lack of knowledge and an attempt to fill the gap with a plausible sounding option: the largest percentage chose the option that defines standard scores as scores obtained with standardized tests. The second has more serious implications. Half the group chose as the acceptable correct answer the statement that mental ability tests measure "underlying capacity for mental functioning." This one error, in our opinion, provides the key to several findings in the rest of our study. The acceptance of this erroneous survival from the early decades of the century would certainly lead to the rejection of certain test uses and to the endorsement of some of the test criticisms found in the preceding section.

The largest percentage of correct answers was given to Item 15, dealing with the meaning of grade-equivalent scores -- whose use, alas, psychometricians have been trying to discourage for many years because of their technical faults. The second largest percentage of correct choices is found on Item 6, which concerns the need for supplementary information in interpreting test scores (somewhat obvious, but gratifying). More

than half the group also chose the correct option, in descending order of frequency, for Items 4 (grade equivalents again!), 5 (age norms), 9 (how norms are established), 7 (what reading readiness tests should measure), 1 (implications of error of measurement and score bands), and 14 (test reliability). The percentage of persons correctly answering the remaining seven items ranged from 17 to 50. The percentage of teachers who marked e, the "Don't Know" option, ranges from 5.9 for Item 6 (need for supplementary information in interpreting test scores) to 49.2 for Item 8 (probability of individual exceeding median of overlapping distribution).[5]

In summary, this sample of widely diversified school teachers demonstrated many gaps in their understanding of basic psychometric and psychological knowledge required for the proper interpretation of test scores. They also revealed evidence of the acceptance of at least one serious popular misconception about what "intelligence" or mental ability tests measure. In view of the fact that all but 10% reported some formal training in educational or psychological testing (Table 1), these findings do not speak well for the effectiveness of their training.

Analysis of Total Scores on Test Opinion Inventory and Test Usage Inventory. In the development of both the Test Opinion

---

[5]From one to six persons marked two options on a few items; these respondents were not credited with a correct response, even if marked, and have been included in the "Don't Know" percentages in Table 6.

Inventory and the Test Usage Inventory, our object was to include items that operationalized a broad-based construct, defined as attitude toward the use of standardized tests in education, in the first case, and knowledge required for proper test interpretation, in the second. We did not seek a high degree of item homogeneity within either inventory. Rather, we endeavored to span a wide diversity of specific item content and situations. By aggregating items heterogeneous in all but the behavior of interest, we can obtain a total score that assesses the defined construct. Whatever empirically established behavioral consistencies remain in this aggregation constitute the trait that is measured. In this regard, we are following the approach recommended by Humphreys (1962, 1970) in the ability domain and more recently by Epstein (1979, 1980) in the personality domain. It is noteworthy that Mischel (1979), an acknowledged champion of situational specifity, recognized the need to operate at different levels of generality for different purposes and to aggregate across the appropriate level of heterogeneity.

While Tables 5 and 6 were concerned with the responses to individual items, Table 7 deals with the characteristics of the total scores on the two inventories, and with the relation between them. On the Test Opinion Inventory, total scores were computed with all item responses adjusted so that high scores represent a

Table 7
Analysis of Total Scores on
Test Opinion Inventory and Test Usage Inventory

Instrument	Total Score Analysis	Correlation between Total Scores
1. Test Opinion Inventory	M 75.40 σ_M=1.14 SD 15.20 Range 23-118 r_{11} .73 N 179	r_{12} = .24 df = 177 p < .01
2. Test Usage Inventory	M 7.74 σ_M=0.19 SD 2.85 Range 0-15 r_{22} .60 N 195	

Note. The reliability coefficients (r_{11} and r_{12}) are measures of internal consistency (coefficient Alpha).

favorable attitude toward tests. The mean score on this inventory is 75.40. This score falls virtually on the 75 midpoint, which would be obtained if all 15 items were given the middle rating of 5. In overall attitude toward testing, this group of teachers can clearly be described as moderate. The range of individual scores, however, is wide, extending from 23 to 118 (out of a possible range of 15 to 135). The internal consistency of the inventory, measured by coefficient Alpha, is .73. This coefficient indicates a substantial level of empirical behavioral consistency across the varied items included in the inventory.

On the Test Usage Inventory, the mean score is 7.74 out of a possible maximum of 15. The individual scores range from 0 to 15. Like the previously discussed analysis of item responses, these total scores suggest serious deficiencies in the background information required for the use and interpretation of test scores. While some individuals were undoubtedly well informed, there are all too many in the group who were not. With regard to internal consistency, the Alpha of .60 again indicates sufficient behavioral consistency for the total score to represent a meaningful construct.

Table 7 also gives the correlation between total scores on the two inventories. This correlation is .24, significant at the .01 level. Although not high, it indicates a clear tendency

for those teachers who are better informed about tests to have more favorable attitudes toward the use of tests in education. This finding supports our hypothesis that adverse attitudes toward testing result in part from misconceptions about test functions and misinterpretation of test scores.

Comparative Analysis of Responses of Black Teachers

In the total teacher sample, the representation of ethnic or cultural minorities other than blacks was too small for meaningful analysis. The entire group includes only three Hispanics, 1 East Indian, one Filipino, and no American Indians or Orientals. With 27 black teachers, the minority sample is in effect a black sample. Thus it was decided to compare the 27 blacks with the total sample. Comparing the black subsample with the total sample, rather than with the white subsample, will of course have a blurring effect, tending to minimize differences. Hence we shall err in the direction of interpretive caution. We wish to avoid focusing attention on small differences, even when statistically significant. It is our object to ascertain merely the extent to which conclusions drawn from the total sample may apply to the black subsample.

Participant Characteristics. A comparison of Tables 1 and 8 reveals certain demographic differences between total sample and black subsample. The proportion of males was substantially

Table 8
Black Teachers: Demographic Data
(N= 27)[a]

Variable	Percentages			
Sex	Female	57.7	Male	42.3
Age	20-29	30.8	50-59	7.7
	30-39	30.8	60 & up	0.
	40-49	30.8		
Education	High School & further	11.1	Master degree	22.2
	Bachelor degree	3.7	Master & further	37.0
	Bachelor & Graduate	25.9	Doctoral degree	0.
Testing Training	Graduate course	40.0	Grad & In-service	4.0
	Undergrad.course	8.0	Undergr.& In-service	8.0
	In-service	12.0	No formal training	16.0
	Grad & Undergr.& In-serv.	12.0		
Reported job title	Classroom teacher	74.1	Specialty teacher (e.g. reading, math.)	7.4
	Special educ.teacher	14.8		
			Principal	3.7

[a] For item responses, N=25-27.

larger among the black teachers, 42% being men as contrasted to only 19% among all the teachers. Age distributions were roughly similar, with the large majority falling between 20 and 49. The black teachers, however, tended to concentrate more heavily in this age range, with very few over 50 and none over 60.

Education shows a somewhat smaller percentage of black teachers with a college degree and a somewhat larger percentage with high school plus additional special training. At the graduate level, however, the proportions are virtually the same. Training in educational or psychological testing showed few differences. Although 16% of the blacks as compared to 10% of the total sample reported no formal training in this area, the proportion reporting a combination of undergraduate, graduate, and in-service training was twice as large among the blacks. Inservice training only was also more commonly reported by the black teachers. The distribution of self-reported job titles among the black teachers is quite similar to that of the total group. The large majority were functioning as classroom teachers, a sizeable group as special education teachers, and a smattering on other school jobs. Thus, apart from the greater proportion of male teachers and the somewhat younger age level, the demographic characteristics of the black sample should not lead to substantial

Table 9
Black Teachers: Student Contacts

Variable	Response Summary			
Direct contact	Yes 100%		No 0%	
Grade levels	Kindergarten to 12			
Number of students	Mdn 14.75		Range 0-250	
Mdn N in ethnic groups	White	0.9	Black	11.3
	Hispanic	0.6	Amer.Indian	0.
	Oriental	0.	Other	0.

differences in our findings.

The data on student contacts, however, suggest some relevant differences (Tables 2 and 9). The entire black group reported direct contact with students. The grade range, K to 12, was virtually the same as for the total group. The median number of students was smaller for the black teachers, a difference that may be related to the greater concentration of black teachers in one or two schools that emphasize individualized instruction. The most conspicuous difference, however, pertains to the ethnic distribution of their own students. The median numbers of white and black students in the total sample were 23 and 2, respectively; among the black teachers, the medians were just under 1 white student and 11 black students. Among individual black teachers, 10 reported having no white students, and an additinnal 6 reported having only one. Only three black teachers reported contacts with large numbers of white students. In contrast, only two black teachers reported having no black students, and the large majority had 10 or more. It would thus seem that the responses of our black subsample may reflect not only the ethnic identity of the respondents but also -- and probably more significantly -- their teaching experience with predominantly black students.

Findings Regarding Test Use. Table 10 gives the percentage of black teachers reporting each test activity; these results

Table 10
Percentage of Black Teachers Reporting Each Test Activity

Activity	Achievement Tests[a]			Mental Ability Tests[b]		
	Regularly	Occasionally	Never	Regularly	Occasionally	Never
Authorize testing program	15.4	26.9	57.7	0.0	19.2	80.8
Select tests	8.0	28.0	64.0	3.8	11.5	84.6
Coordinate, manage program	11.5	26.9	61.5	15.4	7.7	76.9
Decide to continue program	12.5	8.3	79.2	3.7	11.1	85.2
Orient student to tests	34.6	34.6	30.8	11.5	26.9	61.5
Administer tests	46.2	50.0	3.8	22.2	22.2	55.6
Receive test scores	45.8	29.2	25.0	34.6	19.2	46.2
Have access to scores	70.4	11.1	18.5	30.8	26.9	42.3
Use scores: understand students	55.6	37.0	7.4	18.5	40.7	40.7
Use scores: instruction	59.3	33.3	7.4	25.9	33.3	40.7
Use scores: grouping students	37.0	51.9	11.1	18.5	37.0	44.4
Use scores: evaluate performance	37.0	44.4	18.5	14.8	40.7	44.4
Use scores: identify special needs	40.7	48.1	11.1	22.2	37.0	40.7
Counsel students re score	23.1	42.3	34.6	19.2	26.9	53.8
Explain scores: parents or others	16.0	40.0	44.0	16.0	28.0	56.0
Other uses	13.0	21.7	65.2	5.0	20.0	75.0

[a] N=23-27

[b] N=20-27

may be compared with those in Table 3, which gives the corresponding data for the total group. Only a few clear trends emerge from this comparison. With only minor exceptions, the percentage of black teachers regularly performing each activity is larger for achievement than for ability tests, as was found in the total sample. A somewhat larger percentage of black teachers report counseling or instructing students regarding their test scores, and this difference from the total sample is more pronounced for ability than for achievement tests. Using tests in planning and adapting instruction rans higher in relative frequency among black teachers than it does in the total sample, and this difference holds for ability as well as for achievement tests.

The ratings of the usefulness of standardized tests by black teachers coincide closely with those obtained in the total sample (Tables 4 and 11). Again, the mean rating is more favorable for achievement than for ability tests, although the difference is not large. And again, individual differences are wide, the full scale range (1-9) having been used despite the small sample of participants.

Turning to the Test Opinion Inventory (Tables 5 and 12), we find that the black teachers tend to express somewhat more adverse opinions about tests than does the total group. This difference appears both in the ratings of individual items and in the total

Table 11
Black Teacher Ratings of Usefulness of Standardized Tests

Type of Test	N	Mean	SD	Range	σ_M
Achievement	25	5.12	1.97	1-9	.39
Mental Ability	20	4.30	2.27	1-9	.51

scores on the inventory. Reference to Tables 7 and 12 shows a difference of 8.24 in total score, significant at the .01 level. The black group also tends to give more extreme ratings. Despite the small number of cases (and the resulting magnitude of standard errors), nine of the 15 mean ratings deviate significantly from the theoretical neutral value of 5. Two items yield significant deviations in the favorable direction, both at the .01 level. One of these indicates strong rejection of an adverse statement (Item 2, testing is an insult to human dignity); the other indicates strong agreement with a favorable statement (Item 15, proper use of tests prevents unfair discrimination). These two statements were among the four yielding significant deviations in the favorable direction within the total group.

Five items show significant deviations in the unfavorable direction at the .01 level (Items 4, 5, 6, 11, 13) and two at the .05 level (Items 3 and 14). The most extreme ratings represent strong rejection of a favorable statement (Item 5, "IQ" tests are helpful in instructional decisions) and strong agreement with an adverse statement (Item 6, tests tend to underpredict school achievement of minority children). Insofar as school personnel may misinterpret low ability test scores as an indication of "underlying capacity for mental functioning," these attitudes on the part of teachers working with predominantly black groups

Table 12
Black Teacher Responses to Test Opinion Inventory

Item No.	Key Topic	Mean[a]	SD	Range
* 1	Ability tests help instruction.	4.28	2.34	1-9
2	Testing and human dignity.	3.36	2.23	1-9
3	Interviews vs. tests in college admission.	6.00	2.33	1-9
4	Testing is too widespread in our society.	6.68	2.15	3-9
* 5	IQ tests and instructional decisions.	3.23	2.12	1-8
6	Tests underpredict minority school achievement.	7.23	2.50	1-9
* 7	Reading tests aid teachers.	5.64	2.36	1-9
8	Tests measure rote memory.	5.52	2.20	1-9
* 9	Test scores independent of appearance.	5.46	2.37	1-9
* 10	Tests, cultural handicaps, and remediation.	4.00	2.71	1-9
11	Mental ability tests unfair to minorities.	6.52	2.57	1-9
12	Eliminate all standardized tests in education.	4.76	2.15	2-9
13	National norms encourage competition.	6.42	2.21	2-9
14	Measuring only a few traits is harmful.	6.12	2.22	2-9
* 15	Proper test use prevents unfair discrimination.	6.73	2.18	2-9
	Total Score M = 67.76	σ_M = 2.40	SD = 10.46	Range = 49-94

Note. Scale 1-9; low scores indicate favorable attitude, except on starred items. For item responses, N=25-27.

[a] σ_M = .42-.54

of schoolchildren are understandable. Other examples of strongly endorsed test criticisms are found in Item 4(testing is too widespread in our society), 11 (mental ability tests discriminate against minority children), and 13 (national norms encourage competition). It is evident that all these instances of clearly unfavorable attitudes on the part of the black group have special implications for the educational problems encountered by minority children. Apart from the greater prominence of such items, the response pattern of the black teachers is similar to that of the total group. In general, it indicates a thoughtful and meaningful response pattern. While individuals can be characterized in terms of an overall level of favorable-unfavorable attitude toward testing, the manifestations of this attitude are adjusted to the particulars of specific situations.

On the Test Usage Inventory, the black teachers revealed less familiarity with relevant test knowledge than did the total group (Tables 6 and 13). Their mean number of correct responses out of 15 was 5.52, as compared to 7.74 for the total group, a difference that is significant at the .01 level. The range of scores extended from 0 to 10 in this group, and from 0 to 15 in the total group.

Analysis of individual item responses also indicates weaker knowledge of specific test concepts and of relevant background material on the part of the black group. Thus there were several

Table 13
Black Teacher Familiarity with Test Concepts and Interpretations

Item No.	Central Idea	Percentage Choosing Each Response Option				
		a	b	c	d	e
1	Error of measurement and score bands.	25.9	29.6	**18.5**	7.4	18.5
2	Need for norms in interpreting numerical scores.	3.8	11.5	3.8	**53.8**	26.9
3	Nature of criterion-referenced tests.	34.6	0	3.8	**26.9**	34.6
4	Meaning of grade-equivalent scores.	7.7	11.5	**50.0**	7.7	23.1
5	Meaning of age norms.	19.2	**53.8**	0	7.7	19.2
6	Need for supplementary information.	0	16.0	**64.0**	8.0	12.0
7	Nature of reading readiness tests.	34.6	0	**34.6**	3.8	26.9
8	Probability of individual exceeding median of overlapping distribution.	8.3	**20.8**	4.2	12.5	54.2
9	How norms are established.	**46.2**	15.4	7.7	7.7	23.1
10	Environmental handicaps and test scores.	0	**46.2**	23.1	0	30.8
11	Meaning of standard scores.	20.0	20.0	20.0	16.0	24.0
12	What mental ability tests measure	**23.1**	15.4	38.5	7.7	15.4
13	Meaning of percentile scores	24.0	**16.0**	0	32.0	28.0
14	Meaning of test reliability.	**44.0**	4.0	20.0	4.0	28.0
15	Interpretation of grade-equivalent score.	3.8	3.8	3.8	65.4	23.1
Number Correct	$M = 5.52$ $\sigma_M = 0.57$	SD = 2.98		Range = 0-10		

Note. Correct responses are boxed
Response Option e = Don't know --also includes 1 respondent who marked 2 options for item 1.

items in which the percentage of respondents choosing a wrong option was larger than the percentage choosing the correct option (Items 1, 3, 11, 12, 13). Two of these items (11 and 12) were the only two in which a similar reversal occurred in the total sample. In the standard score item (11), the choices of the black teachers reveal a typical chance distribution across the three wrong options, with a smaller percentage choosing the correct option. In the item pertaining to what mental ability tests measure (12), the largest percentage opted for "underlying capacity," thus demonstrating the same misconception found in the total group. As in the total group, also, the largest number of "Don't Know" responses was given for Item 8 (probability of individual exceeding median of overlapping distribution). Similarly, the largest number of correct choices occurred on Item 15 (grade-equivalent scores). The only other items with more than 50% correct responses are Items 2, 5, and 6.

Part II: Testing Coordinators

Procedure

Development of Interview Form. In this part of the study, our principal object was to explore how decisions about school testing programs are made and how tests are selected. For this purpose, we prepared a basic interview schedule which was reviewed ny project consultants specializing in educational

psychology, teacher training, and educational administration at the Fordham University Graduate School of Education. The procedure was also pretested informally on a few persons knowledgeable about testing in the schools of this geographic area.

The interview followed a semi-structured format, with uniform key questions to introduce relevant topics and probing questions for clarification and elaboration. A copy of the interview form is reproduced in Appendix C. Major topics discussed with the respondents include:

(1) Their relative autonomy in decision making
(2) Their budgetary limitations and how this affects their program
(3) Their dependence on year-to-year authorizations and appropriations
(4) The educational decisions for which the test scores are being used
(5) The rationale behind the testing program
(6) The information used in making various testing decisions and where such information is obtained
(7) How respondents evaluate the adequacy of their tests
(8) The frequency with which they have changed tests in the past
(9) How they have attempted to improve teachers' use of test scores
(10) The adequacy of information they receive from test publishers (including what they would like to receive which is now unavailable to them)
(11) The availability and nature of a continuing education program for the teachers in regard to test use

Data-Gathering Procedures. Interviews were held with 15 testing coordinators, drawn from different municipalities or

districts in the tristate metropolitan New York area. The names of the coordinators were obtained from personal contacts of the project staff as well as from some of the testing coordinators themselves. In all cases, a member of the project staff contacted the potential interviewee on the telephone, explained the nature of the project and the interview, and set up an appointment for the interview. A senior member of the project staff then traveled to the coordinator's office so that the interview could be conducted in situ. After the preestablished topics had been covered, the respondents had the opportunity to continue the discussion in any way they wished -- and many of them did so. The interviews varied in duration from 45 minutes to 1½ hours, averaging about 1¼ hours.

We considered the use of a tape recorder, but decided against it after our preliminary explorations because of the adverse effect such a procedure would have on cooperation and rapport. Some schools had had experiences with newspaper reporters who obtained test data through devious means and publicized their findings, with unfortunate after-effects. The result was a heightened sensitivity about any discussion of testing, which we found to be quite general among our participants.

Participant Characteristics

In no case was the participant's official title actually that of "Testing Coordinator". Examples of their specific

job titles include; Director of Research and Testing, Assistant to the Superintendent for Testing and Evaluation, Director of Pupil Assessment and Records, Coordinator of District Testing, Research Associate, Director for Child Placement, and Child Study Team Coordinator. In all 15 cases, the testing coordinators had at least a Master's degree and five had doctorates. All were white; eight were male and seven female.

Perhaps the most conspicuous differences among the participants pertain to their prior job experience and specialized training, and to the routes whereby they reached their present positions. Six had been classroom teachers as well as principals prior to taking this position -- five of these six in the same district in which they were currently working. The sixth had been the principal of a school that closed because of shrinking enrollment, and had subsequently begun working at a neighboring district as testing coordinator. Two of the other former principals had moved into this job within their own districts for the same reason. Three of the testing coordinators had been classroom teachers who applied for their position when it was "posted"; all three explained the rationale behind their movement in terms of increased salary and freedom. Another set of three coordinators had been special education

teachers prior to assuming their coordinator positions. Two of these said that they had become so extensively involved in federal- and state-mandated testing and so knowledgeable about such programs that their districts had asked them to assume their current positions. The third "special educator" was simply desirous of moving into educational administration. One test administrator had worked in the schools, first as a teacher, later as a guidance counselor, and then as an administrator related to guidance, before finally assuming her role as test coordinator.

Two of the testing coordinators had been trained specifically for their positions through graduate work in educational psychology, one with a doctoral degree and one with a master's degree. Neither had worked previously for a state department of education. As will be discussed later, these two coordinators were clearly the best informed with respect to testing, recent developments in the field, and statistical considerations. But they were probably least able to share this knowledge with other school personnel, because their communication networks were relatively undeveloped. One of these respondents, for example, remarked that she was not sure she had ever talked with a teacher in the district about testing! She talked only with administrators, who, in turn, talked with teachers.

With regard to current responsibilities, three coordinators stated that they were in charge of all testing in the district. Ten coordinators were in charge of all standardized group testing, but not individual testing. In seven of these ten cases, their immediate superior was in charge of the school psychologists. In the remaining three, the functions of the school psychologists were coordinated through another office. Finally, two participants reported that they influenced decisions relating to standardized testing, but could not "control" these decisions. On the other hand, these two persons also had responsibilities in relation to individual testing and the placement of specific children within their respective districts.

It is apparent that our 15 testing coordinators exhibit considerable diversity both in their background of experience in the school system and in the nature and extent of responsibilities in their current positions. In these regards, they seem to be quite typical of testing coordinators in the country at large.

Interview Findings

Overview of the Testing Programs. A listing of particular tests used at different grade levels in the districts covered by our respondents would be unwieldy and would jeopardize confidentiality. More important from our standpoint was the coor-

Table 14
Reported Rationales Underlying Test Uses

Rationale	Frequency of Response
Programmatic Reasons	
For funded programs	14
To fulfill state mandate	7
To demonstrate readiness to advance to the the next grade	2
To "flag" poor teachers and help principals	2
To evaluate curricula	2
To assign incoming children to levels	2
Total	29
Reasons Specifically Related to Special Populations	
To diagnose a child's weaknesses for remediation	7
To write a child's Individualized Educational Program (IEP)	4
To decide whether to have child tested by the school psychologist, or consider placement	2
To identify the gifted	2
To justify keeping children in specialized classes	1
Total	16
General Educational Reasons	
To help teachers see their pupils grow	4
To provide information for the yearly parent-teacher conference	3
Total	7

dinator's rationale as to why various tests were being given. The responses to this question are summarized in Table 14. Inspection of this table demonstrates that the most frequent use of standardized tests is to fulfill federal and state guidelines. Of the total 52 respondent comments on this topic, 37 relate to federal and state mandates. This number includes 21 of the 29 programmatic reasons and all 16 of the reasons specifically relating to special populations. Only infrequently were comments made relating to the pedagogical betterment of the schools.

Ability Grouping in the Districts. Of the 15 districts, 12 grouped children according to ability levels and three did not. One of the latter, however, does group children within classes. Of the 12 that grouped children, 11 used test scores, as can be seen in Table 15. Of these 11, four used test scores "almost entirely," four "to some extent," and three "only to a minor extent." In three of the four districts in which tests were used "almost entirely" in grouping children, it was nevertheless reported that teacher judgments were considered at least as collateral information.

Teachers' Use of Test Scores. In 11 of the 15 districts, the teachers received their students' test scores directly. It should be noted, however, that two of the coordinators remarked

Table 15
Sources of Information Used in Grouping Children

Source	Frequency of Response
Test Scores	11
Teacher judgements	11
Parental input	4
Grades	3
Principal's judgement	1
Age[a]	1
Behavior Problems[a]	1

Note. Based on 12 schools that group children.

[a] These comments were made in regard to a "special school."

that teachers received the scores too late to be of se4vice. Of the remaining four districts, three permitted teachers access to the scores, which had been placed in the students' files, and one district blocked teachers from access to their students' scores.

Twelve of the coordinators replied that their teachers were capable of using test scores effectively, while three felt teachers were not able to do so. In addition, 30 replies were received to the question, "In what ways do you think your teachers use test scores?" These are found in Table 16. It It can be seen that the most common uses,in order of frequency, were adapting instruction, aiding individualization, year-to-year comparison of child's performance, and writing Individualized Educational Programs (IEPs).

Only one coordinator commented directly as to why teachers did not use test scores. His explanation was that teachers' jobs are simply too complex already, and that teachers are flooded with so much information that they are not able to spend the time necessary to employ the scores effectively.

We asked a series of questions relating to the coordinator's impressions of teachers' orientations and attitudes about test scores. First, the coordinators were asked whether their teachers felt that test scores were changeable or modifiable, as opposed

Table 16
Teachers' Uses of Test Scores

Reported Uses	Frequency of Response
Adapting instruction	8
Aiding individualization	6
Comparing a child year-to-year	5
Writing IEPs	4
Grouping children (within classes)	1
Teachers do not use test scores	4
Teachers use test scores only if and how their principal tells them to	2

to being fixed or permanent. Of the 15, 12 responsed that teachers did feel the scores were changeable. Among these 12 positive respondents, seven offered additional specific comments. Four said that radical changes in test score were seen frequently--mostly among special-education children who had received remediation. Two commented that they would like to compute change scores to accentuate the amount of change the teachers would see, but that the teachers' unions had prevented the coordinators from doing this. A final respondent explained that teachers were more aware of change in achievement scores than in mental ability scores.

Coordinators were also asked whether they felt the so-called Rosenthal (or Pygmalion) effect was valid. Four responded, "Yes," and eight said, "No." Three were unaware of this effect and, although two of them seemed quite interested in it, all three refused to comment without further information.

We also questioned the coordinators as to whether teachers interpreted test scores, at least in part, from a genetic perspective. Six of the 15 replies were affirmative and nine were negative. Among the six affirmative respondents, three added that administrators perpetuated this approach because it "explained away" why some predominantly minority schools consistently performed poorly on tests. Two additional responses

stated that teachers' unions and teachers individually perpetuated this concept for the same reason. The final affirmative response was coupled with the explanation that teachers saw children from the same families scoring similarly on standardized tests. In this instance, the test coordinator, a former teacher, did not seem cognizant of the fact that these siblings would have had similar family environments as well as common genes.

Finally, the coordinators were asked to describe their impressions of the teachers' opinions about testing, as well as those of administrators, parents, and themselves (as testing coordinators). With respect to teachers, five respondents believed that teachers were very negative about testing. An additional comment referred to teachers' feeling that there was just too much testing presently in the schools. Surprisingly, two independent comments were almost identical: two coordinators stated that two-thirds of the teachers were neutral and one-third were vocally antagonistic. Three other comments stated that teachers had reasonable, professional, well-qualified attitudes about testing, that is, teachers knew that testing was important and necessary and treated it accordingly. Three final comments related to the coordinators' impression that teachers "overused" test scores -- they relied too heavily on test-score information and avoided using other information to

get a "total picture"of the child. One coordinator felt that he was not able to answer the question. Thus, of the 14 comments received, only three could generally be considered positive in regard to the beneficial use of tests by teachers.

With respect to administrator attitudes about testing, 12 coordinators offered opinions, while three replied that they "could not say." Seven stated that administrators were more knowledgeable and positive about the value of testing than were teachers. Two former principals supplemented their positive comments by stating that test-score information helped principals to spend money wisely. Four comments related that principals perceived tests as necessary evils. Another coordinator said (as she had said of the teachers) that two-thirds of administrators were neutral and one-third antagonistic. Thus, seven of the twelve responses, or just over one-half, were positive in regard to good testing use by administrators.

With regard to the coordinators' own attitudes, eight reported that they were more positive about tests than were teachers. The two who felt that teachers relied too heavily on test-score information considered themselves to be more cautious about the benefits of testing than were teachers. Five believed themselves to be about the same as teachers in this regard. In general, however, the interviewer (who had also talked with many of the teachers surveyed in Part I of this study) observed

that the coordinators tended to be much more positive about testing than are typical teachers.

Twelve of the coordinators thought they were unable to comment on parents' attitudes about testing. In most cases, they felt "out of touch," either because they rarely dealt with parents, or because they only dealt with the occasional irate parent and, hence, could not generalize their perception to parents as a whole. Of the three respondents who did comment, two felt that parents lacked the knowledge to properly evaluate tests and their educational benefits. The remaining comment was that parents liked standardized testing: scores helped them to see how their child was doing and permitted them to hold the schools accountable for the child's growth and performance.

Reported Misuses of Tests. The coordinators were asked whether they were aware of any misuses of tests in their district. Even given the promised anonymity of the interviews, it was not surprising that six replied negatively. Of the nine comments made, only one occurred more than once. Most of these "misuses" are not of the classic type leading to incorrect diagnosis or placement of individuals. Two coordinators stated that the local press had "somehow" obtained the test-score means of the schools within the district and had published them,

although these data were supposed to be confidential. The result was somewhat embarassing, insofar as schools differed markedly in mean scores. Another coordinator replied that, against her advice, there were genetic interpretations of test scores, because this interpretation was needed to "explain away" differential school averages and differential school growth rates. Within another district, the coordinator explained that special education teachers received only the overall test scores, whereas subscores are necessary for proper remediation; he felt this was a misuse. One coordinator reported that a secretary at the district level had inadvertently destroyed a number of scores, and that as a result older scores had to be used for administrative purposes. Another coordinator also reported something which he deemed a serious misuse of test scores, and which would, if true, seemingly contradict a considerable body of literature. This coordinator felt that children with high IQs generally receive less than their fair share of the teacher's attention.

Another comment was that, since teachers were evaluated by their students' averages, there had been a case of teachers helping students cheat and providing them with answers in advance of the testing. In another case, the coordinator cited as a misuse the extreme duress that students feel in taking tests. Finally, one coordinator reported that most standardized tests seemed to have vocabulary levels which were too high for his

districts' schoolchildren. Thus, of the nine instances identified by the coordinators as test misuses, two-thirds six related to the interpretation and use of test scores, whereas only two related to test-taking and one to the nature of the tests themselves.

The Job of the Coordinator. Three of the coordinators said that they made most of the testing-related decisions autonomously. Eight made decisions in conjunction with various administrators and other persons. The remaining four described themselves primarily as information providers. They felt that their role was to supply various individuals or groups with relevant information and then letting them make the decisions. The types of persons involved with the test coordinators in making testing-related decisions are given in Table 17. Assistant superintendents and principals lead the list. Four respondents reported Testing Councils, composed of central administrators, curriculum personnel, principals, teachers, and parents.

Testing coordinators were asked to describe how encumbered they were by financial considerations. Nine felt that they did not need any more money. Among these, one said that if he could test less and give back some money to the district, he would. One coordinator did report that he underspent his budget last year and turned back money to the district. Another said that,

Table 17
Types of Persons or Groups With Responsibilities for Making Testing-Related Decisions

Participant Category [a]	Frequency of Response
Assistant Superintendents	6
Principals	5
Testing Councils [b]	4
Teachers	2
Parents	2
Curriculum Personnel	1
School Psychologists	1

[a] In addition to testing coordinators.

[b] Testing Councils are composed of central administrators, curriculum personnel, principals, teachers, and parents.

although finances were adequate, "Every single penny was spent." The respondent from still another district reported that their budget was adequate only because they score their tests themselves.

Six coordinators felt that they needed more money. Of these, four needed it because they wanted teachers to construct more district-level, criterion-referenced tests. One of these also wanted to computerize the test administration process, but was financially unable to do so. Another coordinator wanted money to be able to provide teachers with test-use workshops. The last coordinator said that she could only "make do" on their budget by burying bills on other budget lines not related to testing.

Sources of Test Information. The test coordinators were asked where they went to get information about tests and testing. Their responses are tabulated in Table 18. By far the largest source of information was materials provided by test publishers and their representatives. The footnoted comment in this table is of interest: one coordinator stated that he had attempted using several hardcover texts and had found them all too dated to be of value.

The test coordinators were then asked to evaluate the usefulness and quality of information they received from test publishers. Their responses are found in Table 19. Of the

Table 18
Sources of Information about Tests

Source[a]	Frequency of Response
Publisher Information	13
Consultants	3
Buros MMY	3
Conversations with persons from other districts	2
Back files	2
Notes from class materials	1
State documents	1

[a] One test coordinator stated that hardcover books were invariably out-of-date.

22 responses received, the vast majority were exceedingly negative. In general, the information provided by publishers is seen as biased and not especially helpful, that is, not directed at the questions the coordinators feel require answers. Nonetheless, as stated above, the coordinators use information supplied by test publishers in making testing decisions as their largest source--biased or not.

The coordinators were asked what information, products, and services they would like to receive from publishers that they are not currently receiving. A total of 19 responses were received, with very few repetitions or similar comments. Four coordinators called for quicker scoring services. In connection with computerized score reports, two coordinators requested better individualized diagnostic printouts at reasonable prices, and another two requested better test reports to send home to parents. Two coordinators also suggested the need for better math tests. The remaining responses can be grouped under requests for new services and new materials. With respect to desired services, one test coordinator each asked for the quicker returning of phone call requests for information, in-service training related to the diagnosis of learning disabilities, the identification of other local school districts that are using the same tests (this coordinator actually said that

Table 19
Responses of Coordinators Regarding Test Publisher Information

Statement	Frequency of Response
Is biased and dishonest	9
Does not answer our needs or questions	4
Tries to invoke fear in us (to prevent us from changing tests)	2
Is too oriented to sales, not informative enough	2
Is too hard to understand	2
Must be seen from publishers' perspective	1
Depends on the sales representative; some are good and some are poor	1
Is quite useful	1

two publishers had told him that they had policies against providing these identities), and honest information. With respect to new materials, an additional four comments were made, three of which concerned grade-level problems. One coordinator each requested a listing of items on a test by grade level, readability estimates of tests and test items, and the packaging of various testing materials by grade level. The final comment suggested that more tests be constructed using a task analysis of the measured construct.

The coordinators were next asked how they felt tests themselves should be changed. On the whole, their responses -- even with follow-up questioning -- lacked specificity. Four coordinators, however, called for more district-level tests which are closely tied to the curriculum. Three coordinators each called for better reading tests, tests for the bilingual, and achievement tests that were more behaviorally based and diagnostic. Of the remaining comments, two called for improved testing for special populations (one for the handicapped and one for the learning disabled), and one each called for the Rasch model testing of more skills, the forced external scoring of tests (e.g., outside the district), increased use of task analysis, and reduction of duplication in various tests.

The final set of questions related to the coordinators' feelings about changes in testing that were likely to occur.

Thirty-six responses were made and these are presented in Table 20. By far the largest number of comments predicted a movement away from national norms and toward district-level, curriculum-based tests. In decreasing order of frequency, there was also mention of increased accountability through testing, and the use of equating to reduce duplication in testing.

Conclusions and Reflections. Some general conclusions can be drawn about district-level testing coordinators, even on the basis of our small sample of fifteen. First, the coordinators are generally well-trained professionals, although in most cases they are not in positions for which their specific training and prior experience fitted them. Nevertheless, they are key persons in the sense that they may have access via formal as well as informal communication chains to both high-level district personnel and teachers. They coordinate testing programs which are, in general, adequately funded; these testing programs generally involve the evaluation of governmentally mandated educational activities.

In general, the coordinators believe that tests can have beneficial value to the schools, even though teachers are relatively negative about tests. Principals, they seem to feel, hold the key to whether teachers consider tests as important and use the scores, for example, in planning or adapting in-

Table 20
Comments about the Future of Testing in the Schools

Comments	Frequency of Comment
Movement away from national norms and toward district-level, curriculum-bound tests	13
Increased accountability through testing	5
Use of equating to reduce duplication	3
Better use of test information by teachers and specialists	2
Doing away with labels through the use of IEPs	2
More testing personnel in the schools	2
Better teacher-tester communication	1
Better tests and diagnoses of learning disabilities	1
Better computerized test score reports for parents	1
Less reliance on grade equivalents	1

struction. The coordinators would like to have the time and money to provide workshops on proper test use, although only a minority of them (4) are currently able to hold such workshops.

The coordinators feel trapped. Governmental mandates insist that they employ certain types of tests. At the same time, they are unable to obtain the information that they need about specific tests: standard, impartial sources are too dated, and current test publisher information is too partial and incomplete. The source of information about which several spoke most highly was the use of external consultants not connected with test publishers. Most districts, however, do not appear to have the funds to employ these consultants as needed.

The training of the testing coordinators is critical for the effectiveness with which they handle their positions. Those who have emerged from relatively scientific-psychological backgrounds are less likely to hold serious misconceptions about tests. But they tend to have limited communication patterns with the school system as a whole; rather, they tend to funnel information to decision-makers. On the other hand, former principals generally have had considerable amount of prior experience with the workings of the educational system. And they are able to use this experience in achieving their goals.

It was these former principals, for example, who were most likely to hold workshops for teachers, even though they had less positive attitudes and less knowledge about testing than did the more psychologically trained testing coordinators. Similarly, they were more likely to have the power to make decisions on their own. They were also more distrustful of the test publishers. Improved information directed to these individuals, in the form of written materials, consultation, and workshops, would probably yield the greatest dividends for the improvement of test use in the schools. If these workshops were conducted on a regional basis by the test publishers, they would also probably improve inter-district cross-fertilization and reduce the strong anti-publisher sentiment.

Similarly, the comments about test publishers were so largely negative as to suggest the wisdom of reexamining current practices of test sales. It would probably be fruitful to reduce the sales emphasis and increase the educational function of the publisher representatives. One might even consider paying these representatives totally by salary and avoid commissions; this might reduce the adverse public relations that result from high pressure, impersonal sales. For reasons of confidentiality, test and publisher names have been removed from the statements cited in this report, although they were mentioned frequently in the actual interviews.

There is potential for improving communication with the test coordinators. It must be remembered that these individuals feel loyalty to their school districts rather than to the testing profession. Test coordinators are, in fact, clearly key personnel in that they hold considerable influence throughout the school system. A merging of their local interests with the goals of improved test use should be possible.

PART III: PARENTS

Procedure

Development and Pretesting of Parent Survey. Initially, the parent instrument was developed and pretested as a semi-structured interview. On the basis of pretest data, and because of the availability of PTA groups as participant pools, it was decided to employ small-group procedures with a written questionnaire, thus paralleling the procedures followed with teachers.

Following preparation of a preliminary interview form and initial revisions on the basis of consultant comments, pilot interviews were conducted with a total of 10 parents (8 females and 2 males), all of whom were enrolled in the Fordham University Graduate School of Religion and Religious Education. The interviews were intensive and open-ended, ranging from 30 minutes to two hours. Although the group was small and selected,

especially with regard to educational level, it served its purpose well at the initial stage of instrument development: many responses were obtained to the open-ended questions, and a content analysis of these responses provided useful leads for the subsequent formulation of objective questionnaire items. The pretest results also demonstrated that the questions elicited a diversity of attitudes, the responses of individual interviewees ranging widely for each item -- on a 9-point scale, at least some items yielded a range from 1 to 9. The responses also reflected expected item differences, group means for individual items varying from 3.2 to 7.1.

A preliminary group form of the Parent Survey was developed and pretested in October, 1979 with three new groups of parents drawn from the same source used in pretesting the interview form. The final revision was prepared on the basis of these pretest results. A copy of this form is enclosed in Appendix D.

The Parent Survey consists of a cover page eliciting demographic data about the respondents and their school-age children, a 27-item questionnaire entitled Attitudes Regarding Achievement and Mental Ability Tests, and the same Test Opinion Inventory given to teachers and described in Part I. The "Attitudes Regarding Achievement and Mental Ability Tests," consituting the major part of the Parent Survey, utilizes three

types of items: ratings (on a nine-point scale) of degrees of approval or disapproval of tests or specified uses of tests, checklists for recording both factual and attitudinal responses, and a few open-ended items. The principal topics covered by the items include reactions to the use of tests in general as well as to specific uses of tests, and reactions to one's own children being tested. Feedback regarding children's test performance is explored with reference to whether or not it was received, whether the parent had to request it, the form in which it was received, and how useful the parent found it. All questions are repeated in parallel forms, one dealing with achievement tests and one with mental ability tests.

Data-Gathering Procedures. At an early stage in the project, the decision was reached to call upon PTA groups as participants in this part of our research. If we are interested in ways to improve the use of tests in the schools, we need to begin with the most promising members of the consumer community. Among parents, this means PTA members, particularly those who attend meetings and take an active interest in educational problems. In this sense, our parent samples are selected rather than random. It would certainly be wasteful to begin by trying to communicate with those parents who may lack the time, the experiential background, or the motivation to try to improve educational conditions.

Accordingly, we utilized the same referral sources through which we had obtained teacher groups, namely the faculty and administration of the Fordham University Graduate School of Education and professional contacts available to project staff members. Through these channels, we were able to obtain data from 12 PTA groups -- affiliated with 10 public schools, one parochial school, and one private school -- within the geographical area described in the introductory section of this project. The identical confidentiality commitments were observed with these PTA groups as with the teachers and testing coordinators.

Data gathering procedures followed closely the same steps described for teacher groups at every stage. A total of 223 parents participated, the number included in each session ranging from 8 to 28.

After the completion of data analysis, in October 1980, each PTA group received an individually computerized report of its own data, together with the overall project results on the Parent Survey and an invitation to contact the project staff if further discussion of the report was desired.

Participant Characteristics

Table 21 summarizes the demographic characteristics of the 223 parents who participated in this study. Although the

Table 21
Parent Sample: Demographic Data
(N=223)[a]

Variable	Response Summary			
Sex	Female	92.8%	Male	7.2%
Age	20-29	3.2%	50 - 59	2.3%
	30-39	66.2%	60 & up	0.5%
	40-49	27.9%		
Ethnic	White(non-Hispanic)	94.1%	Hispanic	0.9%
	Black	3.4%	Amer.Indian	0.9%
	Oriental	0.9%	Other	0
Children in School				
Total Number	Mean = 2.17	SD = 1.11		Range=1-8
In Public Schools	Mean=1.81	SD=1.06		Range 0-7
In Parochial Schools	Mean=0.23	SD=0.76		Range 0-4
In Other Private Schools	Mean=1.35	SD=0.44		Range 0-3
Ages	Mean = 10.7	SD = 5.7		Range = 4-18+

[a] For item responses, N = 220-222

group consisted chiefly of females, 7% were males. In age, the group fell predominantly between 30 and 39, with slightly over a fourth in the 40 to 49 decade, and very small percentages at younger and older age levels. All had children in school, the mean number being slightly over 2 and the range from 1 to 8. The largest number of children were attending public schools, with private and parochial schools following in that order. The children ranged in age from 4 years to 18 years, with a mean age of 11 years.

Total Group Results

Rated Uses of Standardized Tests. In the first section of the Parent Survey, respondents recorded their opinion of various uses of achievement and ability tests on nine-point rating scales, ranging from strongly disapprove through neutral to strongly approve. The identical uses were rated for achievement and mental ability tests on successive scales. Table 22 gives the means and standard deviations for each rating. The first pair of scales called for an overall rating of the use of the two types of tests with schoolchildren. The next five pairs were concerned with the use of tests in making specific educational decisions about individuals, including tracking, giving extra instruction, influencing career plans, and identifying

Table 22
Parent Ratings of Uses of Standardized Tests

Test Use Rated	Achievement Tests		Mental Ability Tests	
	Mean[a]	SD	Mean[b]	SD
Overall use in schools	6.04	2.29	5.45	2.42
Tracking decisions	5.37	2.61	4.80	2.59
Additional instruction	7.43	1.89	6.37	2.34
Influence career plans	4.38	2.55	4.07	2.44
Identify intellectually gifted	5.61	2.69	5.41	2.77
Identify children with learning disabilities	5.49	2.75	5.25	2.70

Note. Scale 1-9; obtained range 1-9 for each item.
High scores indicate favorable attitude.
For item responses, N = 219-223.

[a] σ_M = .13-.19

[b] σ_M = .16-.19

intellectually gifted students and children with learning disabilities.

It should be noted, first, that the mean ratings are consistently higher for achievement than for ability tests. Although varying in amount, this difference holds across all uses. There is complete consistency, however, in the rank order of these ratings for achievement and ability tests. In this regard, the type of decision to be made on the basis of tests seems to weigh more heavily than does the type of test employed. The highest ratings are assigned to the use of tests as a basis for giving students added instruction. The mean ratings for this use of testing, 7.43 for achievement tests and 6.37 for ability tests, deviate by large and statistically significant amounts from the theoretical neutral value of 5.

The overall rating, which comes next in rank order, is also significantly favorable for both types of tests. The uses of tests for identifying gifted and learning disabled children, in that order, are also rated favorably. Tracking decisions are rated only slightly above 5 with regard to achievement tests, but fall below 5 with regard to ability tests. Career decisions based on either type of test tend to be disapproved, both mean ratings falling significantly below the neutral value at the .01 level. It is noteworthy that both tracking and career decisions have long-term implications. One

senses here a suspiciousness about possible irreversible decisions and the labeling of children on the basis of improperly interpreted test scores.

Test Uses in Own Child's School. Parents were next asked to consider the same five types of test uses discussed in the preceding section, this time with reference to current practice in their own child's school. Our object in this set of questions was to assess how much parents know about the way tests are used in their children's schools and how they perceive such uses. Table 23 reports the percentage of parents who reported that tests were or were not used for making decisions in the specified areas, as well as the percentage who stated they did not know. The same questions were asked with reference to achievement tests and mental ability tests.

An examination of Table 23 shows that, as perceived by this group of parents, achievement tests were used much more extensively than were ability tests for all types of decisions. Despite the diversity of purposes listed, the percentages of respondents indicating that achievement tests are employed for tracking, additional instruction, and the identification of gifted and learning disabled children are closely similar, ranging from 60 to 67. Influencing career plans stands out as the sole exception, with only 10%. The role that early test results may

Table 23
Percentage of Parents Reporting Test Uses in Own Child's School

Test Use	Achievement Tests[a]			Mental Ability Tests[b]		
	Yes	No	Don't know	Yes	No	Don't know
Tracking decisions	60.0	10.7	29.3	32.3	23.3	44.4
Additional instruction	67.0	7.9	25.1	32.1	20.2	47.7
Influence career plans	10.3	31.5	58.2	7.9	33.2	58.9
Identify gifted children	63.3	9.8	27.0	45.3	18.9	35.8
Identify children with learning disabilities	61.9	7.8	30.3	43.2	16.8	40.0
Other	37.7	-	-	27.4	-	-

[a] N = 213-218

[b] N = 189-193

play in channeling interests and steering individuals into different educational programs is probably too subtle to be readily apparent. In fact, over half of the parents reported that they did not know whether achievement or ability tests were used for this purpose in their child's school.

Turning specifically to mental ability tests, we find somewhat greater differentiation among the different test uses. The more distinctly "educational" decisions, such as tracking and the provision of extra instruction, are less often influenced by ability tests than are the more "psychological" decisions that call for more causal analysis and understanding, such as the identification of gifted children and of students with special learning disabilities.

In two open-ended questions, parents were asked to list any other uses of achievement and ability tests that they believed were being made in their child's school. The responses mentioned essentially the same uses for both types of tests. Several parents merely repeated uses already listed, often specifying more detail or providing concrete illustrations: the most common examples referred to tracking and the identification of specific disabilities. There was some mention of accountability, with reference to both teachers and schools. The most numerous set of new comments pertained to understanding and evaluating the child's performance in terms of both cognitive

and noncognitive variables. In this connection, some questionable test uses were cited, such as labeling children and concluding that a child is lazy if he or she performs more poorly on achievement than on ability tests. The old stereotypes and misconceptions about these two categories of tests are certainly found among parents, as they are among teachers. And of course, the teachers may have communicated such interpretations to the parents.

Finally, we may examine the percentage of "Don't Know" responses. Apart from the previously mentioned "career plans" item, which yielded 58% and 59% such responses for achievement and ability tests, respectively, the proportion is consistently and substantially larger for mental ability tests. The elimination of group "intelligence" tests from several of the schools we surveyed may account in part for this discrepancy. On the other hand, when mental ability tests are administered on an individual basis by professional examiners, there is more need to discuss the findings with the child's parents. Whatever the reason, the percentages of parents reporting that they did not know whether tests were used for the remaining four types of decisions ranged from 25 to 30 for achievement tests, and from 35 to 48 for ability tests. Also relevant to these findings is the relatively large number of nonrespondents in this portion of the questionnaire, ranging from 5 to 10 for achievement tests

and from 30 to 34 for ability tests. We can assume that most of these nonrespondents also lacked knowledge about current testing practices in their child's school. On the whole the indications are that this group of concerned, active parents, attending PTA meetings, was not as well informed about the use of tests in their children's schools as would be desirable.

Reactions to Test Use with Own Child. In exploring parents' reactions to the use of tests with their own children, we followed two approaches. First, on a pair of nine-point rating scales, parents indicated how they felt about their child being tested in school with achievement and mental ability tests, respectively. The mean ratings, given in Table 24 are 6.03 for achievement tests and 5.56 for ability tests. Both indicate approval beyond the indifference point of 5, the deviations being significant at the .01 level in both cases. The approval is slightly stronger for achievement than for ability tests, but the difference just falls short of significance at the .05 level.

Approaching the question from a second angle, we asked parents whether they would prefer that their children receive more or less testing in school. The replies are summarized in Table 24 for achievement and ability tests. The largest percentages -- 48 for achievement and 35 for ability tests -- reported satisfaction with the present amount of testing. The

Table 24
Parent Reactions to Use of Tests with Own Child

Question	Response Summary	
	Achievement Tests	Mental Ability Tests
Approve of child being tested[a]	M=6.03 SD=2.50 N=223 σ_M= .17	M=5.56 SD=2.65 N=218 σ_M= .18
Percentage preferring:		
More testing	6.4	9.0
Less testing	30.6	26.1
Present amount	48.4	35.2
not sure	14.6	29.6
	(N=219)	(N=199)

[a] Scale = 1-9, obtained range = 1-9
High scores indicate approval

next largest percentages (31 and 26) wanted less testing; and the smalles percentages (6 and 9), wanted more testing. A sizeable percentage indicated that they were not sure, this percentage being twice as large for ability as for achievement tests (30 vs. 15). To this finding may be added the presence of 24 nonrespondents for ability tests as compared to only 4 for achievement tests.

Immediately following the question in which parents were to check that they preferred more or less testing, were satisfied with present amount, or not sure, they were asked, "Why do you feel this way?" Among the parents wanting more testing, the most frequent reason given for both achievement and ability tests was that such testing provides feedback to parents and follow-up information. Additional reasons listed for more achievement testing included that it increases the child's motivation, sharpens test-taking skills, and contributes to accountability of teachers and schools.

The principal reasons for preferring less testing of both types (in decreasing order of joint frequency) include the opinions that tests are: probably not accurate; generally not useful; too time-consuming and/or expensive; too emotionally stressful; used in making unwarranted judgments and labeling children; not used by teachers to follow-up on scores or make

comparisons over time; and culturally or racially biased. It should be noted that the last reason was listed by only one respondent for achievement tests and only two for ability tests. One comment contributed by six respondents with reference to achievement tests only was that teacher judgment of class work provides a better estimate of student performance than do achievement tests.

The parents who were satisfied with the present amount of testing generally gave reasons that were similar to those cited in support of more testing. Quite understandably, those who were not sure gave a combination of reasons that others had cited in support of more or less testing. For example, they might say that testing is valuable but emotionally stressful. Other respondents mentioned only an advantage or only a disadvantage of testing as the reason for their indecision. The one new comment contributed by an "unsure" but thoughtful parent was that ability tests are useful for special children only.

Feedback Regarding Test Results. Because of our concern regarding communication, we were especially interested in the nature and extent of feedback these parents received regarding the test performance of their own children. The results are summarized in Table 25. Of the respondents, 81% reported that they had received such information about achievement tests

Table 25
Percentage of Parents Reporting Kinds of Feedback Regarding Child's Tests

Kind of Feedback	Achievement Tests (N=217)	Mental Ability Tests (N=192)
Information received	Yes: 80.6 No: 18.4 Unsure: 0.9	Yes: 39.1 No:54.7 Unsure: 6.3
Had to ask	44.9	61.2
Without asking	55.1	38.8
Form of transmittal:[a]		
Individual conference-teacher	33.1	33.3
Individual conference-other	16.0	35.6
Parent meeting	3.0	3.3
Teacher letter	4.5	1.1
Other school letter	2.2	2.2
Score report-no explanation	8.9	7.8
Score report-with explanation	26.4	12.2
Teacher message through child	1.5	1.1
Conversation with child	1.9	1.1
Other	2.6	2.2

[a] Percentages based on total number of responses given; some parents reported more than one form of transmittal.

and 39% that they had received it about ability tests. There was generally less certainty about ability tests: 6% of the respondents indicated that they were not sure whether or not they had received ability test feedback, and 31 parents failed to answer this question. It is quite likely, of course, that in most of these cases no mental ability test had been administered to the particular children.

The remaining percentages in Table 25 are based only on respondents who had received information on the appropriate type of test. Of these, nearly half (45%) said they had to ask for the information in the case of achievement tests, and well over half (61%) said so in the case of ability tests. The next question pertains to the form in which the feedback was transmitted. The percentages refer to the total frequency of each form of transmittal reported; some respondents had received information in more than one way.

About a third of the communications regarding both types of tests were received in an individual conference with a teacher. For ability tests, a slightly larger percentage (36) were transmitted in a personal conference with another school official. Only 16% of achievement test communications were received in this manner. A score report with appended explanation was the next most frequent mode of communication, accounting for 26% of the transmittals for achievement tests and 12% of those for

ability tests. Score reports without explanations accounted for 9% and 8%, respectively. Other forms of communication were relatively infrequent. It might be noted that only 3% were received at a parents' meetings, where the nature of the tests and the interpretation of scores could be explained to the group, while written summaries of the individual results would be distributed to parents.

Although it is gratifying to find that the more desirable means of transmittal predominated, the appreciable frequency of other, less communicative procedures is disturbing. This situation is highlighted by some of the free responses contributed by parents in connection with other, unlisted means of transmittal. Seven parents had received the feedback from a clerk, secretary, or school office worker[6]; five had received it on the report card or attached to it; two received it in a telephone conversation, and one from a teacher she had met in the supermarket.

Parents were also asked to indicate how <u>useful</u> they had found the feedback they had received. This reaction was recorded on a nine-point rating scale ranging from "completely useless" (1) through "fairly useful" (5) to "extremely useful" (9). The means, reported in Table 26 indicate that the group

[6]Six were from a single school.

Table 26
Parent Ratings of Effect of Own Child's Test Taking

Effect	Achievement Tests			Mental Ability Tests		
	N	Mean	SD	N	Mean	SD
Usefulness of feedback	177	5.00 σ_M=.18	2.41	94	4.81 σ_M=.28	2.74
Effect of test-taking experience	191	5.07 σ_M=.13	1.72	124	5.23 σ_M=.14	1.53

Note. Scale 1-9, obtained range 1-9 for each item.
High scores indicate favorable attitude.

found achievement test information fairly useful, while ability test information is rated less useful by a small and statistically insignificant amount. Nevertheless, we must note that the ratings given by individual parents ranged from 1 to 9.

In an open-ended question, respondents were asked what they believed the school or teacher might have done to inform them better about their child's test performance. Three types of suggestions accounted for nearly all the responses: 43 parents recommended increased feedback, primarily through conferences to explain the child's strengths and weaknesses; 27 asked for explanations about the purpose of the test and its implications for the child; and 15 suggested that parents should be automatically informed about test results rather than having to ask for them.

Finally, parents were asked to evaluate the effect that the test-taking experience had had on their own child. The two rating scales used for achievement and ability tests ranged from "quite harmful" (1), through "no effect" (5) to "quite helpful" (9). The mean ratings for both achievement and mental ability tests fall virtually on the "no effect" rating (5.07 and 5.23). It should nevertheless be noted that the ratings assigned by individual parents again ranged over the entire scale, from 1 to 9.

Test Opinion Inventory. The same inventory used to assess the attitudes of teachers toward testing was administered as the last section of the Parent Survey. Table 27 gives the means and standard deviations of the ratings assigned to each statement. While the entire scale range (1-9) was utilized for each item by individual respondents, most of the group means cluster around the scale midpoint, suggesting a moderate attitude. Since the standard errors of the obtained item means range from .16 to .19, a conservative estimate of a statistically significant deviation ($p < .01$) from the theoretical value of 5 is indicated by differences of approximately ± 0.5 or higher. Five of the items have means that do not differ significantly from 5; these are Items 6, 8, 9, 10, and 13.

Let us now consider the ten items whose means do deviate significantly from the neutral value. Of these, half demonstrate a favorable attitude toward tests, three through rejection of adverse statements and two through acceptance of favorable statements. In order of magnitude of deviation, the rejected adverse statements include Items 2 (testing is an insult to human dignity), 12 (elimination of all standardized testing in educational system), and 11 (ability tests discriminate against minorities). The two showing strong acceptance of favorable statements are Items 15 (properly used tests prevent unfair dis-

Table 27
Parent Responses to Test Opinion Inventory

Item No.	Key Topic	Mean[a]	SD
* 1	Ability tests help instruction.	4.14	2.48
2	Testing and human dignity.	3.47	2.46
3	Interviews vs. tests in college admission.	6.14	2.30
4	Testing too widespread in our society.	5.71	2.55
* 5	IQ tests and instructional decisions.	4.44	2.49
* 6	Tests underpredict minority school achievement.	5.14	2.41
* 7	Reading tests aid teachers.	5.81	2.53
8	Tests measure rote memory.	4.85	2.32
* 9	Test scores independent of appearance.	5.12	2.48
* 10	Tests, cultural handicaps, and remediation.	4.72	2.44
11	Mental ability tests unfair to minorities.	4.39	2.60
12	Eliminate all standardized tests in education.	3.79	2.54
13	National norms encourage competition.	5.03	2.74
14	Measuring only a few traits is harmful.	5.55	2.51
* 15	Proper test use prevents unfair discrimination.	5.93	2.38

Total Score M = 75.85 σ_M = 1.53 SD = 20.82 Range = 17.27 r_{11} = .85

Note. Scale 1-9, obtained range 1-9 for all items.
Low scores indicate favorable attitude except on starred items.
For item responses N = 209-218

[a] σ_M = .16-.19

crimination) and 7 (standardized reading tests are helpful to teachers).

An equal number of items yield significant deviations in the unfavorable direction. Of these, three represent acceptance of adverse statements. They include, in order of magnitude of deviation, Items 3 (interviews should receive more weight that tests in college admission), 4 (testing is too widespread in our society), and 14 (measuring only a few traits is harmful). Two represent rejection of favorable statements, namely Items 1 (ability tests help instruction) and 5 ("IQ" tests aid instructional decisions).

In general, the response pattern of these parents shows a reasonable and thoughtful differentiation between the more extreme and the more moderate statements, whether favorable or unfavorable. There is some indication of a more unfavorable attitude toward ability than toward achievement tests, illustrated by Items 1, 5, and 7, and a concern with the proper use of tests (Item 15) rather than a blanket objection to all standardized testing (Item 12). There is also evidence that the group as a whole does not regard standardized tests as such to be unfair to minorities (Items 11 and 15).

The total inventory scores yield an internal consistency coefficient (alpha) of .85. There is thus sufficient interitem behavioral consistency to justify the use of total scores as an

index of a meaningful attitudinal construct. The mean total score is 75.85, which coincides closely with the neutral or moderate value of .75. The range of individual differences is wide, extending from 17 to 127. The extreme score of 17, which falls close to the possible minimum of 15, was obtained by a single atypical respondent, the next higher score being 41.

Relation between Testing Feedback and Parental Attitudes. One of the hypotheses we wished to test regarding communication was that parents who received feedback about their children's test performance without having to ask for it would have more favorable attitudes toward testing than would those who received such feedback only after asking for it. Table 28 summarizes the data relevant to this hypothesis. The indices of parental attitudes (dependent variables) examined for this purpose include the ratings on overall use of tests with schoolchildren (Table 22), the ratings on having one's own child tested (Table 24), and total scores on the 15-item Test Opinion Survey (Table 27).

It will be seen in Table 28 that, with regard to feedback about achievement tests, none of the three differences reaches statistical significance, although all three are in the hypothesized direction. With mental ability tests, however, the hypothesis is supported at the .01 level on the first question, and at the .05 level on the second question and the inventory scores.

Table 28
Relation between Testing Feedback
and Parental Attitudes on Testing

Attitude Measure		Without Asking	Had to Ask	t-ratio of Mean Differences[a]
Feedback about Achievement Tests				
Overall Use of Tests in Schools	M	6.23	5.75	1.40
	SD	2.28	2.32	df 176
	N	98	80	n.s.
Approval of Own Child Being Tested	M	614	6.03	.31
	SD	2.56	2.50	df 176
	N	98	80	n.s.
Test Opinion Inventory: Total Score[b]	M	75.30	74.67	.18
	SD	22.12	19.26	df 143
	N	82	63	n.s.
Feedback about Mental Ability Tests				
Overall Use of Tests in Schools	M	6.27	4.98	2.44
	SD	2.55	2.27	df 83
	N	33	52	$p < .01$
Approval of Own Child Being Tested	M	6.44	5.24	2.23
	SD	2.68	2.20	df 81
	N	32	51	$p < .05$
Test Opinion Inventory: Total Score[b]	M	81.75	72.91	1.71
	SD	24.85	18.80	df 70
	N	28	44	$p < .05$

[a] one-tailed test.

[b] All items scored with favorable ratings hign; possible range 15-135, neutral score 75.

The contrast between the achievement and ability test results is even more conspicuous when we recognize that the number of parents who received information about ability tests, with or without having to ask, is much smaller than the number receiving achievement test information. Hence large differences were required to reach statistical significance with the ability test results. This contrast in results is consistent with the finding throughout our study that the distinction between achievement and ability tests is very real for parents as well as for school personnel. Regardless of the minimal intrinsic differences in the nature of the two types of tests, they are used in different ways, largely by different persons, and in different situations. And they are differently perceived by parents and teachers.

With reference specifically to the attitude findings reported in Table 28, we can speculate that achievement test feedback was perceived by parents in a more routine fashion, as an adjunct to regular school instruction. Thus the form of communication and the conditions under which it was received had little impact on parental attitudes toward testing. Feedback on mental ability tests, on the other hand, was in a different category. These test scores have more emotional significance and ego involvement; they are more likely to be misinterpreted and misused; and at least in some of the schools surveyed, they were

not routinely administered in group sessions, but were given to individual children for special reasons by a school psychologist or other professional specialist. When results of such tests are regularly communicated by the school on its own initiative, they are more likely to be adequately explained. As a result, the parents will be more receptive to the need for such testing, the use to be made of the results, and the interpretation of the scores. If the parent had to ask for such feedback, on the other hand, he or she probably learned that an ability test had been given to the child and may have decided to ask for the results because of apprehension generated by that knowledge. The kind of feedback received under these conditions may also have been less adequate and less reassuring, because the school may not have had effective established mechanisms for transmitting such information.

The causal relation between requesting feedback and attitude toward testing could, of course, have operated in the reverse direction in some cases. Parents who are suspicious of tests or hostile toward their use may be more likely to ask for feedback when their child is tested, so that they may detect anticipated misuses and misinterpretations and may complain about them. But this attitude itself is likely to have developed as a result of cumulative past experience regarding tests. Adequate and correct communication is certainly an effective means of pre-

venting such unfavorable buildup and of minimizing its carry-over to subsequent test use.

Comparative Analysis of Responses of Black Parents

The sample of 223 parents who participated in this study included only 7 blacks, 2 Hispanics, 2 Orientals, and 2 American Indians. It is apparent that minority populations were not represented in sufficient numbers to justify separate analyses. Nevertheless, we decided to examine the responses of the seven black parents, from the viewpoint of descriptive rather than inferential statistics. Because of greater demands on their time, black parents as a whole are probably less likely to participate in PTA activities than are white parents. Consequently, those who do participate are probably a more highly selected sample with regard to interest in educational matters and knowledge about educational problems. It was thus felt that the responses of our small group of concerned and knowledgeable black parents were of interest in their own right.

Participant Characteristics. The black parents included six females and one male. Age distribution is similar to that of the total sample, with four persons between 30 and 39, one in the 20s, and 2 in the 40s. The group had a total of 10 children in school, ranging in age from 4 to 17, with a mean age of 10 years. Nine attended public schools, one a parochial school.

Attitudes Regarding Standardized Tests. Our analysis con-

centrates on those parts of the Parent Survey in which attitudes toward testing were explored through various approaches. These data fit into a meaningful unit that provides an opportunity for internal aggregation and corroboration among the various responses. Because of the small number of cases, it was not considered productive to analyze responses to isolated items dealing with different questions about test use. All the findings from our analysis of black parent attitudes have been assembled into two tables, in which the responses of the black group are compared with those of the total parent sample.

Table 29 summarizes the ratings obtained in response to various items in different parts of the inventory. Looking first at the comparative evaluation of test uses, we find the mean ratings by black parents to be fairly similar to those from the entire sample. For achievement tests, the most conspicuous differences are the somewhat stronger disapproval of using these tests to influence career decisions and the stronger approval of their use for identifying the intellectually gifted. For ability tests, the black parents express somewhat stronger approval of the overall use of these tests, as well as of their use for giving extra instruction and identifying learning disabilities. They express somewhat stronger opposition to the use of ability ttsts for tracking decisions.

Approval of having one's own child tested shows little

Table 29
Comparative Ratings of Standardized Tests by Black Parents

Question	Achievement Tests					Mental Ability Tests				
	Total[a]		Black			Total[a]		Black		
	M	SD	M	SD	Range	M	SD	M	SD	Range
Test Uses:										
Overall use in schools	6.04	2.29	6.29	2.22	3-9	5.45	2.42	6.29	2.98	2-9
Tracking decisions	5.37	2.61	5.86	3.08	1-9	3.80	2.59	4.00	2.83	1-8
Extra Instruction	7.43	1.89	7.86	1.46	6-9	6.37	2.34	7.00	2.31	3-9
Influence career plans	4.38	2.55	3.43	2.88	1-8	4.07	2.44	3.71	2.63	1-7
Intellectually gifted	5.61	2.69	7.00	2.00	4-9	5.41	2.77	5.71	2.43	2-9
Learning disabilities	5.49	2.75	5.14	2.73	1-8	5.25	2.70	6.00	2.38	3-9
Approve of own child being tested	6.03	2.50	5.71	3.40	1-9	5.56	2.65	5.57	2.94	1-9
Usefulness of feedback	5.00	2.41	6.50	1.70	4-8	4.81	2.74	6.00[b]	-	-
Effect of test-taking experience	5.07	1.72	6.60	1.14	5-8	5.23	1.53	7.50	2.12	6-9

Note. Scale 1-9, high scores indicate favorable attitude

[a] Range 1-9 for all items in Total Group

[b] Response by a single person, who gave a rating of 6

difference between blacks and the total group. Both usefulness of the testing feedback and effect of the test-taking experience on their children are given more favorable ratings by the black parents, and this difference holds for both achievement and ability tests.

Table 30 provides a comparison of the responses of the black parents and the total sample on the Test Opinion Inventory. We note, first, that this small group of black parents tends to express more extreme attitudes. Their item means range from 1.67 to 7.60, while in the total group the range is from 3.47 to 6.14. Moreover, in the black group, seven items receive mean ratings of 6 or over, in contrast to only one item in the total group. The distribution of high and low ratings assigned to items favorable and unfavorable to testing, however, is such as to indicate no overall attitude difference between the black group and the total group. This is corroborated by the total inventory scores, in which the black mean (75.60) is virtually identical with the total group mean (75.85). It will be recalled, furthermore, that this mean corresponds to the theoretically defined neutral value of 75 (15x5). These findings suggest that, even more than the total group, these black parents expressed clearly differentiated views, rather than an overall bias for or against testing. Their responses tend to show disapproval of real or potential misuses of tests, while favoring proper and constructive uses.

Table 30
Comparative Responses of Black Parents to Test Opinion Inventory

Item No.	Key Topic	Total		Black		
		M	SD	M	SD	Range
* 1	Ability tests help instruction.	4.14	2.48	4.17	3.13	1-9
2	Testing and human dignity.	3.47	2.46	2.83	3.13	1-9
3	Interviews vs. tests in college admission.	6.14	2.30	6.17	2.71	1-9
4	Testing too widespread in our society.	5.71	2.55	5.16	3.92	1-9
* 5	IQ tests and instructional decisions.	4.44	2.49	5.50	3.21	1-9
6	Tests underpredict minority school achievement.	5.14	2.41	6.50	2.95	1-9
* 7	Reading tests aid teachers.	5.81	2.53	5.83	2.04	3-9
8	Tests measure rote memory.	4.85	2.32	5.83	2.71	3-9
* 9	Test scores independent of appearance.	5.12	2.48	6.00	2.00	4-9
* 10	Tests, cultural handicaps, and remediation.	4.72	2.44	5.50	3.02	2-9
11	Mental ability tests unfair to minorities.	4.39	2.60	6.50	3.33	1-9
12	Eliminate all standardized tests in education.	3.79	2.54	1.67	1.03	1-3
13	National norms encourage competition.	5.03	2.74	7.20	1.48	5-9
14	Measuring only a few traits is harmful	5.55	2.51	7.00	2.10	4-9
* 15	Proper test use prevents unfair discrimination.	5.93	2.38	7.60	2.61	3-9
	Total Score	M=75.85 SD=20.82	Range=17-127	M=75.60 SD=20.48	Range=48-100	

Note. Low item scores favorable, but scale reversed on starred items and in total score.

[a] All ranges 1-9.

An examination of the individual items yielding means that deviate by one point or more from the theoretical value of 5 illustrates this response pattern. Of the 15 inventory items, nine yield means at or beyond these limits. Four of these items indicate a favorable attitude toward tests. These include two very strong rejections of adverse statements, which characterize testing as an insult to human dignity (Item 2) and advocate the elimination of all standardized testing in the educational system (Item 12).[7] The other two responses represent strong acceptance of favorable statements, referring to the impartiality of tests which are unaffected by appearance and manner (Item 9), and to the function of properly used tests in preventing unfair discrimination (Item 15). It is noteworthy that both of these statements recognize the potential value of tests for minority groups. Both show stronger acceptance by the blacks than by the total group.

Five responses represent unfavorable opinions regarding tests. All of these indicate acceptance of adverse statements. In order of magnitude of deviation, they include Items 13 (standardized tests encourage competition with others rather than self-improvement), 14 (tests too limited in ability coverage), 6 (tests underpredict school achievement of minorities), 11 (mental

[7] Despite the small number of cases, this item-mean deviates significantly from 5 at $p < .01$, owing to the narrow range and low SD.

ability tests discriminate against minorities), and 3 (give more weight to interviews than to test scores in college admission). The last of these (Item 3) yields virtually the same mean as in the total group; the other four show stronger agreement with the stated criticism among the blacks. In all these statements, one can see evidence of possible past experience with the kinds of misuses and misinterpretations of tests than can have more adverse effects on minority than on majority students.

Comparative Analysis of Parent and Teacher Responses on Test Opinion Inventory

The Test Opinion Inventory was administered in the identical form to all parent and teacher samples in our study. Table 31 brings together the results obtained in both groups, including item responses and total scores. To begin with the total scores, we find virtual identity in group means, both falling practically on the theoretical midpoint of the scale. The parents, however, exhibit wider individual differences, with an SD of 20.82, as compared with 15.20 for the teachers.

With regard to individual items, the response pattern is predominantly similar for the two groups. Following our previously established rule-of-thumb that means below 4.5 and above 5.5 represent significant deviations from the neutral 5, we find four items showing significant favorable deviations among the

Table 31

Comparison of Parent and Teacher Responses on Test Opinion Inventory

Item No.	Key Topic	Parents			Teachers		
		Mean	SD	σ_M	Mean	SD	σ_M
* 1	Ability tests help instruction.	4.14	2.48	.17	4.43	2.20	.16
2	Testing and human dignity.	3.47	2.46	.17	3.22	2.13	.15
3	Interviews vs.tests in college admission.	6.14	2.30	.17	6.26	2.06	.14
4	Testing too widespread in our society.	5.71	2.55	.18	6.28	2.13	.15
* 5	IQ tests and instructional decisions.	4.44	2.49	.17	4.48	2.15	.15
6	Tests underpredict minority school achievement.	5.14	2.41	.16	6.28	2.20	.16
* 7	Reading tests aid teachers.	5.81	2.53	.17	6.13	1.95	.14
8	Tests measure rote memory .	4.85	2.32	.16	4.62	2.11	.15
* 9	Test scores independent of appearance.	5.12	2.48	.17	4.83	2.32	.16
*10	Tests, cultural handicaps, and remediation.	4.72	2.44	.17	4.30	2.42	.17
11	Mental ability tests unfair to minorities.	4.39	2.60	.18	5.59	2.52	.18
12	Eliminate all standardized tests in education.	3.79	2.54	.17	3.62	2.35	.16
13	National norms encourage competition.	5.03	2.74	.19	5.05	2.52	.18
14	Measuring only a few traits is harmful.	5.55	2.51	.17	5.33	2.42	.17
*15	Proper test use prevents unfair discrimination.	5.93	2.38	.16	6.37	2.00	.14
	Total Score	M = 75.85	SD = 20.82		M = 75.40	SD = 15.20	
		σ_M = 1.53	Range=17-127		σ_M = 1.14	Range = 23-118	

Note. Scale 1-9, ranges 1-9 for each item in both groups. Low scores favorable, except on starred items.

teachers (Items 2, 7, 12, and 15). The parents' means yield favorable deviations on the same four items, plus one other, Item 11 (mental ability tests are unfair to minorities). While the parents reject this item (4.39), the teachers accept it (5.59), the group difference being significant at $p < .01$.

A total of seven item means indicate significant unfavorable attitudes for the teachers (Items 1, 3, 4, 5, 6, 10, 11). Four of these items show a deviation in the same direction for parents. The three items yielding inconsistent group responses include the previously cited Item 11, which deviated in opposite directions in the two groups, and Items 6 and 10, with significant unfavorable deviations among the teachers but no significant deviation from neutrality among the parents. Both of these items deal with the use of tests with minority groups. The Item 10 mean is significantly different between the two groups; the Item 6 mean is not. In addition, one item mean (Item 14) just barely falls into the significant unfavorable range for parents, while remaining in the neutral range for teachers. The group difference in this item does not itself reach statistical significance, however, and the finding is probably negligible.

This rapid overview of item responses on the Test Opinion Inventory indicates that teachers tended to be more critical of tests than were parents in certain specific situations, all

involving the use of tests with minorities. We could speculate that the teachers have witnessed and been involved in more misuses of tests with minority children than have the parents and are more often responding in terms of these misuses rather than in terms of potential constructive uses.

DISCUSSION

Overview

Our specific findings regarding the current use of tests in the schools have been summarized in the three major sections of this report. In those sections, the factual results were presented in relation to the particular questions we set out to answer with each of our participant groups: teachers, testing coordinators, and parents of schoolchildren.

In our study of 207 teachers, we wanted to know what functions the teachers themselves perform in testing; how they use test scores in their teaching; how useful they consider standardized achievement and ability tests, and how they think their usefulness can be improved; what are their attitudes toward the use of standardized tests in the educational system in general; to what extent they have acquired the background knowledge needed for the proper interpretation of test results; and what is the relation between their mastery of this body of knowledge and their attitude toward tests.

In the interviews with 15 testing coordinators, we were principally interested in learning on what bases and by whom decisons about tests and testing programs are made. At the same time, we were able to obtain clarifying and supporting information on several questions raised in other parts of the study. The data we collected from the testing coordinators occupy a special place in our study, for at least two reasons. The first pertains to the intrinsic features of the interviewing technique, including the richness of detail it can provide and the opportunity it affords to coordinate and integrate the discrete facts obtained from any one respondent. Second, and equally important, is the key position occupied by testing coordinators, not only in influencing testing decisions but also in transmitting knowledge and attitudes to the grass-roots users of test results in the schools.

In our survey of 223 parents, we investigated their attitudes toward different uses of standardized achievement and ability tests in the schools; how much they know about the ways in which tests are actually used in their own child's school; how they feel about their own child being tested, and whether they would prefer an increase or a decrease in such testing; whether they received feedback about their child's test performance, and if so, whether they had to ask for it or received it without asking; how the feedback was transmitted to them, and how useful

they found it; whether they believe testing has had a helpful or harmful effect on their child; their attitudes toward the use of standardized testing in the educational system in general; and whether such attitudes show any relation to the conditions of feedback transmittal.

The answers to these specific questions can be found in Parts I, II, and II of this report. Further condensation and summary of the results would be redundant. In the present section, therefore, we will highlight certain broad conclusions of special significance. These conclusions are supported by findings from different parts of the study, or from different questions asked within any one part.

Communication in Relation to Standardized Testing and Test Results

The aura of confusion and malaise that surrounds contemporary testing has multiple causes and hence calls for multiple remedies. In this study, we have concentrated on one cause, which we consider to be among the most important. That cause centers on communication, broadly conceived and operating on several levels. At the simplest and perhaps most direct level, there is need for improving feedback about test results to teachers, school administrators, and parents.

At a more fundamental level, there is need for wider and more effective dissemination of testing knowledge to test users. This calls for accurate and up-to-date information about the

nature and function of tests, the meaning of scores on standardized tests, and the findings of behavioral science that are relevant to the proper interpretation and application of test results. This type of communiction requires different forms of transmittal and different content and language to adapt it to the needs of the general public, parents of schoolchildren, teachers, and other educational personnel involved with testing in various capacities. There is ample evidence from a variety of sources tapped in our study that the current means of communication in this area have not proved effective with any of the publics to which they are directed.

At a third level, we need to keep open the channels of communication from teachers and parents to school administrators and to test publishers and authors. Because of the inadequacies of communications about testing to teachers and parents, misconceptions and misuses are still prevalent. Augmented communication from these sources can help to identify such misuses and can in turn identify the kinds of information about testing whose transmittal requires strengthening.

The Special Needs of Teachers

That teachers as a group do not have sufficient prerequisite information about testing for the proper interpretation of test scores was indicated by the results of the Test Usage Inventory.

This is a 15-item multiple-choice inventory designed to sample basic testing concepts and types of scores, as well as the interpretive context of test results. Both the teachers' total scores on this inventory and their responses to individual items revealed serious gaps in essential knowledge. The mean total score corresponded to barely over fifty percent correct responses. Analysis of the percentage of teachers choosing each response option in individual items revealed confusions and incorrect information about several basic psychometric concepts. Even more disturbing is the finding that one of the items dealing with a fundamental interpretive orientation yielded more choices for an incorrect option than for the correct option. To the question, "In general, what do mental ability tests measure?" the majority chose the response option, "underlying capacity for mental functioning," in preference to "current level of intellectual performance." From another angle, the sharp interpretive differentiation between so-called achievement tests and so-called mental ability tests that still pervades the school system contributes to the perpetuation of outworn stereotypes about the functions of testing.

The acceptance of popular misconceptions, as well as the paucity of correct information about psychometric concepts, is especially noteworthy when we consider that the group of teachers

participating in this study had a relatively high level of education, the largest proportion reporting a Master's degree plus further graduate training. They also had been exposed to considerable specialized training in testing; the largest proportion had completed a graduate testing course, and a sizeable proportion had received various combinations of in-service training and academic courses in testing. It thus appears that the existing training facilities in the schools of education and in-service courses cannot be relied upon to do the whole job of preparing teachers adequately for the proper interpretation and use of test results.

It may also be noted that test knowledge, as assessed by the Test Usage Inventory, was significantly correlated with overall attitude toward standardized testing, as assessed by the Test Opinion Inventory. While not large, this correlation shows a clearly established tendency for teachers who are more knowledgeable about testing concepts and interpretations to express more favorable attitudes toward testing.

The black teachers in our sample, who also proved to be the teachers with the largest number of black students, performed more poorly on the Test Usage Inventory than did the total sample. They also exhibited a significantly more critical attitude toward testing on the Test Opinion Inventory, especially in those items dealing with the applications of testing to minority children.

This more highly negative attitude may itself be associated with their more frequent acceptance of popular misconceptions about tests. It may also reflect the misuses and misconceptions they have witnessed in the interpretations others have attributed to the test performance of minority children.

It should be added that, when asked what can be done to improve testing practices, the suggestions offered by some teachers, as well as some testing coordinators, called for ways of improving teachers' knowledge about tests. Examples included workshops on testing, fuller explanations about the nature and functions of particular tests, and more aids and guides to interpretation accompanying test-score feedback.

Lack of time is a second major obstacle to the full and effective use of test scores by teachers. When some testing coordinators expressed the opinion that teachers do not make use of test scores in their teaching, they were asked why this is so. The answer was, in effect, that teachers already have too much to do. Testing and the availability of test scores merely add to their burdens. Insofar as this condition prevails, it highlights the need to simplify the teacher's task of applying the information provided by tests to their day-by-day classroom activities. This is being done increasingly in the case of diagnostic achievement tests, through detailed response analyses and associated instructional prescriptions provided in the test

manuals, and through elaborated score reports. Also indicated is the need for professionally trained personnel to bridge the gap between test results and their instructional implications in individual cases. Some of the teachers' suggestions for the improvement of test utilization referred to additional professionally trained persons in the schools and to professionally oriented and more knowledgeable publishers' representatives.

A third major problem surfaced in the course of our interviews with testing coordinators. Several respondents observed that genetic explanations of individual differences in test performance are often encouraged by teachers' organizations, school administrators, and individual teachers because they "explain away" school differences in mean scores and in rate of progress over time. Teachers certainly need protection against improper and unsound accountability procedures. Superficially, it appears that mean test scores and measures of score changes over time provide a readily accessible and objective way of assessing the effectiveness of teachers. Such a procedure, however, ignores the powerful influence on student performance and progress exerted by home and family, peer groups, and the community at large. And it ignores the contributions of antecedent experience to the student's readiness to learn at any academic level. Actually, the proper use of student test performance as an

index of accountability for evaluating teachers, instructional programs, or schools requires sophisticated experimental designs in order to establish causal relations and to allocate effects to the appropriate variables (McDonald, 1976).

Substituting a genetic explanation in place of a proper understanding of the modifiability of developed abilities is a way of seeking protection from an unsound accountability practice by promulgating a fallacy. It is a fragile defense, and one that creates other problems, such as the rejection of tests. If one begins by assuming that tests are supposed to measure innate potential, then, when it is demonstrated that scores reflect individual differences in prior experience -- as can certainly be done -- the tests are branded as invalid and biased.

The Concerns of Parents

Basically, parents are concerned about possible misuses of tests, which may jeopardize their child's subsequent educational progress. This concern was manifested in various ways in answers to our questions. Parents expressed more satisfaction with achievement than with ability tests, as did teachers. This reaction may reveal their good judgment in preferring the more forthright, reasonable, and objective interpretations of performance traditionally associated with achievement tests to those traditionally associated with ability tests. If one must live in the 1920s, as some educators and even a few test constructors seem

to be doing, then indeed achievement tests are more acceptable than are tests of academic aptitude, intelligence, educational potential, mental ability, or any other tests with squishy labels.

In their evaluation of different uses of tests, our group of parents evidenced thoughtful discrimination rather than a generalized positive or negative bias toward testing. For instance, the use of tests for deciding which children should receive additional instruction was rated high. Testing to identify gifted children and children with learning disabilities also received favorable mean ratings. The use of tests for tracking and their use in influencing career decisions, however, tended to be disapproved.

The parents' responses in the Test Opinion Inventory were equally discriminating. Of the 15 items in this inventory, 10 yielded mean responses indicating a significant deviation from a neutral attitude, 5 in a favorable, and 5 in an unfavorable direction. In general, favorable attitudes were expressed by strong rejection of extreme test criticisms and strong acceptance of statements implying proper test uses. The unfavorable reactions indicated objection to the purely routine use of test scores, and a recognition of the need for other data in reaching decisions. They also revealed a suspicion of tests labeled "IQ" and those with the more innocuous but not sufficiently reassuring "mental

ability" label.

Our survey also indicated that these parent groups were not, on the whole well-informed about the way tests are actually being used in their children's schools. From a fourth to more than a half did not know whether or not tests were used in making various kinds of educational decisions. When asked whether they approved of their child's being tested, the mean response was on the favorable side of the scale. When asked about the amount of testing, the largest percentage expressed satisfaction with the current amount. Among those who wanted more testing, the most frequent reason given was that it provides feedback to parents and follow-up information. It might be noted that we also heard practically the same statement from those testing coordinators who gave us their impressions about parents' attitudes toward testing. Among the principal reasons given by those parents who wanted less testing was that tests are used in making unwarranted judgments and in labeling children -- and presumably this kind of label is attached with long-lasting glue.

Viewing parental information from another angle, we explored the kind of feedback parents received about their children's test performance. A sizeable percentage reported they had not received testing feedback or were uncertain about it. Among those who had received it, nearly half reported that they did so only after having

asked for it. It is interesting to note that, on the Test Opinion Inventory, those who had received information without having to ask for it expressed a significantly more favorable attitude toward tests than did those who had to ask for it.

The methods of transmittal of the test-score information varied considerably. The more desirable methods -- those providing some opportunity for discussion or explanation -- were the most frequent. Nevertheless, several routine and less communicative procedures were still followed all too often. This situation was reflected in the parents' comments about ways to make the feedback more useful to them. The suggestions given by nearly all respondents centered on more explanatory feedback, including a discussion of the child's strengths and weaknesses, as well as information about the nature and purpose of the test. Also recommended was the regular receipt of feedback without having to request it.

Although the number of black parents in our study was too small for drawing generalizable conclusions, the reactions of this small sample are of interest in that they corroborate the findings of recent polls (e.g., Gallup, 1979), as well as the views expressed by many black educational leaders. Essentially, our results suggest that these black parents were cognizant of the value of properly used standardized tests in furthering their

children's educational progress and in combating racial bias.

On the whole, the findings of our study indicate that parents' misgivings about the use of tests in the schools could be met by improved communication at two levels; first, better feedback about their children's test results; and second, more dissemination of information about the nature and functions of standardized tests. The relatively concerned and knowledgeable parents who attend PTA meetings, from whom our parent samples were drawn, represent a receptive audience for communications about testing. We draw this conclusion, not only from the responses we received in our survey, but also from informal indications of a strong interest in testing discussions, as illustrated by invitations for a return visit by project staff and by requests that a copy of our findings be sent to the local board of education.

The School in a Societal Context

Some of the problems about test use, including the stubborn and viable misconceptions about the meaning of test scores, derive in part from societal pressures. These pressures reflect the fact that schools are embedded in the larger society which includes government bureaucracies, the news media, teachers' associations, parent organizations, and a multiplicity of special interest groups. Hints about the operation of these pressures surfaced

from many sources in our study, but they were especially evident in the responses of the testing coordinators.

The pressures from different groups interact in complex ways. For example, the media stir up and publicize a controversy by focusing on statements by a few, atypical writers, such as Arthur Jensen. To give the controversy more news value, they encourage the false impression that these are the views generally held by psychometricians and psychologists. With further distortion and simplification, what emerges is the notion that all intelligence tests are designed to reveal underlying capacities that are stable, enduring, and unmodifiable throughout the individual's life span. Parents become alarmed by these manifestly absurd claims, which can so readily be disproved. Minority group members, whose experiential background may differ in many ways from that of the population on whom the tests were standardized, become especially concerned about the implications of using such instruments.

Beginning with these highly publicized misconceptions about what mental ability tests are supposed to do, it is easy to deduce most of the objections to testing that have been expressed. It is also easy to deduce that the objections will be stronger on the part of persons with atypical experiential backgrounds, stronger against "mental ability" than against "achievement" tests, and

stronger against standardized tests with national norms than against locally constructed, nonnormative, or so-called criterion-referenced tests.

Another example of interrelated societal pressures was mentioned in connection with our discussion of the needs of teachers. Parents, as well as taxpayers who support public education, demand accountability from the schools. Partly from test results and partly from other observations, people have become aware of a general decline in basic skills and other aspects of academic achievement. Some try to hold teachers responsible for the fact that students do not learn as well as they should. In an effort to defend themselves against this accusation, some teachers and administrators, in turn, fall back on a genetic explanation of slow learning, thereby accepting the interpretation that the media are both publicizing and attacking. Such a defense on the part of teachers also introduces another misconception, namely that genetically derived characteristics are fixed and unmodifiable, while environmentally derived characteristics are evanescent, temporary, and readily susceptible to brief and limited interventions.

Where does all this leave us with regard to the criticisms that have been directed against standardized tests? First, insofar as these tests are occasionally misused and misinterpreted

in the schools, the criticisms are justified. This situation calls for the transmission of more and better information about testing to school personnel. Second, even when tests are properly used, the prevailing negative attitudes toward tests may be generalized to include all test uses. Prior experience with real or alleged misuses may lead to the assumption that tests will be generally misused to the detriment of children. In this case, more effective communications from schools to parents about test uses with their own children should help to relieve some anxieties. The situation will also benefit from wider dissemination of general testing knowledge to parent groups and to the general public.

The Role of Test Publishers

The larger society in which the schools are embedded contains test publishers. When our testing coordinators were asked where they obtained the information they needed to make testing decisions, the large majority replied that they rely principally on information provided by test publishers. Yet these same coordinators expressed consistent dissatisfaction and distrust regarding the information they received from this source. Several specific examples of these reactions were cited in our summary of results from the study of testing coordinators. In our teacher survey, the respondents were asked what could be done to make the tests more useful to them. For both achievement and ability tests, the

most frequent responses pertained to the improvement of feedback from test publishers. Specific suggestions included faster return of test scores; further breakdown of scores; and the use of simpler, clearer language in explanations directed to teachers, parents, and students.

It must be borne in mind, of course, that both testing coordinators and teachers had been exposed to personal and written communications from many different publishers. There is undoubtedly wide variation among test publishers with regard to the specific criticisms expressed by the participants in our study. It should also be recognized that the reported comments indicate the perceptions of the respondents, rather than direct observation of the sources of these perceptions. But, of course, it is these perceptions that influence attitudes and behavior.

With thse caveats, we must conclude that all test publishers could profitably reexamine their whole system of communication with schools. This includes written communications, ranging from test-score reports to test manuals, brochures describing services and products, and special publications designed to improve test users' understanding of tests and the interpretation of test scores. It also includes personal communications in connection with individual telephone calls and written inquiries. It certainly includes the training and orientation of field

representatives. And it includes the possibility of sponsoring workshops on test interpretation, conducted for key school personnel by persons who are themselves adequately informed.

In the various parts of this discussion, we have illustrated how the findings of our study of teachers, testing coordinators, and parents support the communication hypothesis. Breakdowns and inadequacies of communication may lead to misuses of tests, as well as to criticisms of tests, both justified and unjustified. And improved communication offers a promising means of meeting these problems.

APPENDIX

A. Sample Letters Regarding Confidentiality Commitment

B. School Personnel Inventory

C. Interview Schedule for Testing Coordinators

D. Parent Survey

Appendix A

Sample Letters Regarding Confidentiality Commitment

Graduate School of Arts and Sciences

FORDHAM UNIVERSITY Bronx, N.Y. 10458

Department of Psychology

February 26, 1980

Ms. Jane Doe, Principal
The Exemplary Elementary School
15 Education Road, South
Shangri-La, New York

Dear Ms. Doe:

I wanted to write and tell you how pleasant it was to meet with you on the fifteenth of this month. I contacted Mr. Blank yesterday and we plan to talk this coming Friday. I wanted to assure Dr. Zilch in writing that in no way will any individuals in your school, your school itself, the school district, town, county, etc. ever be identified in any way. You will have complete anonymity. On the other hand, we feel that when the teachers complete the questionnaires, they will find the questions responsive to their feelings and not at all directive. We assure you that the data we collect may well have an impact on the educational community and that we will be able to give you some useful feedback with respect to your own school. I look forward to seeing you again on March 10 at 3:00 (or shortly before 3:00).

Sincerely yours,

Kurt F. Geisinger, Ph.D.
Assistant Professor
Project Director

KFG/pm

c.c. Dr. John Zilch

Graduate School of Arts and Sciences

FORDHAM UNIVERSITY Bronx, N.Y. 10458

Department of Psychology

April 25, 1980

Dr. John Roe
Assistant Director of Special Education
Utopia County
Erewhon, New York

Dear Dr. Roe:

I would like this letter to confirm our recent phone conversation. We are conducting a research project concerning teachers' use of standardized tests, as well as their attitudes toward these tests. I have talked with Ms. Jane Blank with respect to surveying her teachers at a faculty meeting and she was quite amenable. In return, we will provide Ms. Blank with feedback with respect to the responses of her faculty. Also, I promise you, as I did her, that no mention of any of the individuals in the study, of the schools in question, or their location, proximate or general, will be made. I am happy to thank you for your cooperation in this important investigation.

Sincerely yours,

Kurt F. Geisinger, Ph.D.
Assistant Professor
Project Direct

KFG/pm

APPENDIX B

SCHOOL PERSONNEL INVENTORY

BACKGROUND INFORMATION

Instructions: Please respond to the following questions, either by checking the appropriate space or by writing in the answer.

1. Sex: Male____ Female____

2. Age: 20-29____ 50-59____
30-39____ 60 & up____
40-49____

3. Ethnic group:
White (other than Hispanic)____ Hispanic____
Black____ Oriental____ American Indian____
Other (please specify)_____

4. Your present educational background (check highest appropriate level):
High school + further training____
Bachelor's degree____
Bachelor's degree + graduate training____
Master's degree____
Master's degree + further graduate training____
Doctoral degree____

5. Your training in educational and/or psychological testing (check one or more spaces):
Graduate level course____ In-service training____
Undergraduate-level course____ No formal training____

6. Your current job title: _____

_____

7. Do you work directly with students? Yes____ No____
If yes, answer 8, 9, and 10 below.

8. With what grade level(s) of students do you work? _____

_____

9. What is the total number of students with whom you have classroom or other professional contact? _____

10. Of these, approximately how many are members of each of the following ethnic groups:
White (other than Hispanic)____ Hispanic____
Black____ Oriental____ American Indian____
Other (please specify)_____

TEST ACTIVITIES INVENTORY

PART I: ACHIEVEMENT TESTS

Please circle the frequency of your participation in your present capacity in the activities listed below. Answer with respect to standardized, achievement, or subject matter tests. Standardized tests are uniform, widely used tests as distinguished from tests made by teachers for use in their own classes. Achievement tests measure how well children have learned subject matter taught in school; examples include achievement batteries, tests of basic skills in reading or mathematics, tests in history, etc. Respond to each activity if you either perform the activity yourself or serve as a member of a committee performing that activity. Remember, this part concerns achievement tests; you will be asked the same questions about mental ability tests in Part II on page 4.

Activity	Frequency (Circle one for each activity)		
Authorize the introduction of the testing program in your school(s)	Regularly	Occasionally	Never
Select a test from several tests for use at your school(s)	Regularly	Occasionally	Never
Coordinate and manage the testing program	Regularly	Occasionally	Never
Decide on the continued employment of the testing program at your school(s)	Regularly	Occasionally	Never
Orient students to take tests	Regularly	Occasionally	Never
Administer tests to students	Regularly	Occasionally	Never
Receive the test scores	Regularly	Occasionally	Never
Have access to these test scores (when not received directly)	Regularly	Occasionally	Never
Use test scores in helping to understand students	Regularly	Occasionally	Never
Use test scores in planning and adapting instruction	Regularly	Occasionally	Never
Use test scores for grouping children for instructional purposes	Regularly	Occasionally	Never
Use test scores to evaluate performance (of individual students, classes, programs, and instruction)	Regularly	Occasionally	Never
Use test scores to identify children with special needs	Regularly	Occasionally	Never
Counsel or instruct students regarding the meaning of their test scores	Regularly	Occasionally	Never

Continue on the next page

Activity	Frequency		
Explain the meaning of these test scores to parents or to teachers or other school personnel	Regularly	Occasionally	Never
Use the test scores in some other manner (please specify below)	Regularly	Occasionally	Never

_____

_____

_____

1. In general, how useful have you found standardized achievement tests in your work? Circle the number which corresponds to your evaluation of their usefulness

Not at all Useful		Mildly Useful		Moderately Useful		Quite Useful		Highly Useful
1	2	3	4	5	6	7	8	9

2. What could be done to make these tests more useful to you?

_____

_____

_____

_____

_____

_____

3. Please name one or more specific examples of standardized subject matter tests with which you have recently been involved.

_____

_____

_____

TEST ACTIVITIES INVENTORY

PART II: MENTAL ABILITY TESTS

Part II of this inventory deals with the standardized mental ability tests. Mental ability tests are often referred to as tests of intelligence, general aptitude, academic potential, or school readiness. The questions below are the same as those you answered in PART I. Please circle the frequency of your participation in your present capacity in the activities listed below. Respond to each activity if you either perform the activity yourself or serve as a member of a committee performing that activity.

Activity	Frequency (Circle one for each activity)		
Authorize the introduction of the testing program in your school(s)	Regularly	Occasionally	Never
Select a test from several tests for use at your school(s)	Regularly	Occasionally	Never
Coordinate and manage the testing program	Regularly	Occasionally	Never
Decide on the continued employment of the testing program at your school(s)	Regularly	Occasionally	Never
Orient students to take tests	Regularly	Occasionally	Never
Administer tests to students	Regularly	Occasionally	Never
Receive the test scores	Regularly	Occasionally	Never
Have access to these test scores (when not received directly)	Regularly	Occasionally	Never
Use test scores in helping to understand students	Regularly	Occasionally	Never
Use test scores in planning and adapting instruction	Regularly	Occasionally	Never
Use test scores for grouping children for instructional purposes	Regularly	Occasionally	Never
Use test scores to evaluate performance (of individual students, classes, programs, and instruction)	Regularly	Occasionally	Never
Use test scores to identify children with special needs	Regularly	Occasionally	Never

Continue on next page

Activity	Frequency		
Counsel or instruct students regarding the meaning of their test scores	Regularly	Occasionally	Never
Explain the meaning of these test scores to parents or to teachers or other school personnel	Regularly	Occasionally	Never
Use the test scores in some other manner (please specify below)	Regularly	Occasionally	Never

_____

_____

_____

1. In general, how useful have you found standardized mental ability tests in your work? Circle the number which corresponds to your evaluation of their usefulness.

Not at all Useful		Mildly Useful		Moderately Useful		Quite Useful		Highly Useful
1	2	3	4	5	6	7	8	9

2. What could be done to make these tests more useful to you?

_____

_____

_____

_____

_____

_____

3. Please name one or more specific examples of standardized mental ability tests with which you have recently been involved.

_____

_____

_____

TEST OPINION INVENTORY

Please circle the number which corresponds to the extent to which you agree or disagree with the statements below.

	Strongly Disagree				?				Strongly Agree
1. Giving mental ability ("intelligence") tests would improve the effectiveness of instruction.	1	2	3	4	5	6	7	8	9
2. The attempt to measure human behavior (i.e. attitudes, achievement, aptitude, personality, etc.) is an insult to human dignity.	1	2	3	4	5	6	7	8	9
3. In admitting students, colleges should give more weight to personal interviews than to test scores.	1	2	3	4	5	6	7	8	9
4. Testing is too widespread in our society.	1	2	3	4	5	6	7	8	9
5. "I.Q." tests are quite helpful in making instructional decisions about students.	1	2	3	4	5	6	7	8	9
6. Tests designed to predict how well children will do in school tend to underestimate the school achievement of minority children.	1	2	3	4	5	6	7	8	9
7. Scores on standardized reading tests are an important aid to teachers.	1	2	3	4	5	6	7	8	9
8. Most achievement tests assess mainly rote memory for unimportant details rather than true understanding.	1	2	3	4	5	6	7	8	9
9. Tests can assess impartially a child's knowledge and reveal talent, because they are not affected by the child's appearance, clothing, or manners.	1	2	3	4	5	6	7	8	9

	Strongly Disagree				?				Strongly Agree
10. Tests help to identify specific cultural handicaps as a necessary first step in planning remedial action.	1	2	3	4	5	6	7	8	9
11. Most tests of "mental" or "scholastic" ability have been built in such a way as to discriminate against minority children.	1	2	3	4	5	6	7	8	9
12. The educational system would be better off if all standardized testing were eliminated.	1	2	3	4	5	6	7	8	9
13. Standardized tests with national norms are undesirable because they encourage comparison and competition with others instead of focusing on the improvement of one's own performance.	1	2	3	4	5	6	7	8	9
14. There are so many abilities and characteristics relevant to school performance that assessing only one or two of these characteristics does more harm than good.	1	2	3	4	5	6	7	8	9
15. When properly used, tests can serve an important function in preventing irrelevant and unfair discrimination.	1	2	3	4	5	6	7	8	9

Test Usage Inventory

Instructions: The questions below are concerned with how test scores are actually used and interpreted in dealing with schoolchildren. Please respond to every question by circling your preferred answer. If for any question you are not sure which answer to choose, please circle"e"(Don't Know).

1. On a certain test, it was found that a child's score is likely to change by a few points when the child is retested at different times or with different forms of the test. The best way to handle this situation is to:
   a. retest and report the second score.
   b. retest and report the higher score.
   c. test once and report a likely range for each child.
   d. ignore the test because it is unreliable and cannot be trusted.
   e. Don't Know

2. Sally, a fourth-grade child, earned a score of 98 on a nationally standardized, mental aptitude test. What is the best interpretation of this score?
   a. Sally completed nearly all items correctly.
   b. Sally is approximately average in general intelligence.
   c. Sally has the capacity currently to perform far beyond the typical fourth-grade level.
   d. Sally's score is meaningless until we know the test norms.
   e. Don't Know

3. A school district has developed a criterion-referenced reading test. Which of the following would best represent the fact that the test is valuable?
   a. The test-retest reliability of the test is high.
   b. A vast majority of the students pass the test.
   c. A significant portion of the students fail the test.
   d. The test accurately reflects reading competence.
   e. Don't Know

4. Grade-equivalent test scores can best be described as:
   a. scores which assign a letter grade (A, B, C, D, F) to the child's performance.
   b. scores which correspond to letter grades earned in school.
   c. scores which compare individual test scores to norms for each grade.
   d. scores which are based on the school grade in which the child is currently enrolled.
   e. Don't Know

5. On a 100-item aptitude test, the average number correct for 10-year-olds is 85. The number 85 can be best designated as a:
   a. mode
   b. norm
   c. standard score
   d. percentile score
   e. Don't Know

6. A student is not achieving well in school. She is given an intelligence test and earns an extremely high IQ. Which of the following best describes your interpretation of the situation and your possible actions?
   a. There is probably something wrong with the intelligence test and she should be given another test.
   b. She is an underachiever. Her high aptitude should be discussed with her and she should be urged to work harder.
   c. More information about her should be gathered in the attempt to understand why she is not achieving well.
   d. She probably has an emotional problem and should be referred to a school psychologist.
   e. Don't Know

7. Of the following measures, which would you consider most useful to assess the reading readiness of a first-grade child?
   a. a culture-free test that reveals innate ability unaffected by past experience
   b. a test of practical problem solving
   c. a test covering prerequisite skills identified through an analysis of the reading process
   d. a test of basic attitudes and motivation
   e. Don't Know

8. If 30 percent of black children reach or exceed the median of white children on an intelligence test, the chance out of 100 that a black child picked at random will score at or above the white median is:
   a. 15.
   b. 30.
   c. 50.
   d. 70.
   e. Don't Know

9. In standardized ability tests, norms are most often established by:
   a. testing a representative group.
   b. conducting controlled experiments.
   c. applying psychological theory.
   d. obtaining consensus of experts.
   e. Don't Know

10. If an intellectual difficulty can be attributed to environmental handicaps in a child's background, this difficulty:
   a. should be discounted because it is not part of his true nature.
   b. should be reflected in test scores so that remedial procedures may be applied.
   c. should not affect the child's score on a properly constructed culture-fair test.
   d. will be outgrown after the child is placed in a normal school environment.
   e. Don't Know

11. Standard scores can be most accurately described as test scores:
    a. obtained under standard testing conditions.
    b. obtained with standardized tests.
    c. evaluated in terms of an objective standard.
    d. expressed in terms of standard deviation units.
    e. Don't Know

12. In general, what do mental ability tests measure?
    a. current level of intellectual performance
    b. test-taking abilities and skills
    c. underlying capacity for mental functioning
    d. the speed of manipulating abstract concepts
    e. Don't Know

13. If Debby receives a percentile score of 70 on a standardized achievement test, which of the following conclusions is more likely to be true?
    a. She answered 70% of the items correctly.
    b. Her performance was clearly better than average.
    c. Her performance was clearly poorer than average.
    d. Her performance was just about average.
    e. Don't Know

14. The term "reliability" refers primarily to which of the following characteristics of test scores?
    a. consistency
    b. objectivity
    c. accuracy
    d. variability
    e. Don't Know

15. If a boy earned a grade-equivalent score of 3.5 on a reading test, it means that:
    a. he reads about average for children of his age.
    b. he had three questions correct and one partially correct.
    c. his performance rates a B+.
    d. he reads about as well as an average mid-year third-grader.
    e. Don't Know

	Strongly Disagree				?				Strongly Agree
10. Tests help to identify specific cultural handicaps as a necessary first step in planning remedial action.	1	2	3	4	5	6	7	8	9
11. Most tests of "mental" or "scholastic" ability have been built in such a way as to discriminate against minority children.	1	2	3	4	5	6	7	8	9
12. The educational system would be better off if all standardized testing were eliminated.	1	2	3	4	5	6	7	8	9
13. Standardized tests with national norms are undesirable because they encourage comparison and competition with others instead of focusing on the improvement of one's own performance.	1	2	3	4	5	6	7	8	9
14. There are so many abilities and characteristics relevant to school performance that assessing only one or two of these characteristics does more harm than good.	1	2	3	4	5	6	7	8	9
15. When properly used, tests can serve an important function in preventing irrelevant and unfair discrimination.	1	2	3	4	5	6	7	8	9

APPENDIX C

INTERVIEW SCHEDULE FOR TESTING COORDINATORS

Interview Schedule for Testing Coordinators

Introductory Orientation Statement: Explain that we are interested in finding out (1) what their testing program is, (2) why it has been developed the way it was and how he/she makes decisions related to the testing program, (3) how tests are used in the school district, and (4) how he/she relates to and receives information from people outside the district (e.g. test publishers, research centers, the state and federal governments). Explain that the overall goal of this research effort is to describe school test programs. Remind him/her that neither he/she nor the school district will be identified. Entertain questions.

Description of the program: Ask him/her to describe their testing program. (What grade levels are tested with what tests?) What other testing goes on in the district (e.g., by guidance counselors, school psychologists, the state, etc.)? What are his/her relationships to these personnel/organizations?

The Rationale of the Program: Ask for what purpose each test in this program is being given (e.g., Title I evaluation, evaluation of teachers, evaluation of students' progress, ability grouping, etc.)

Grouping of Children: Query the coordinator as to whether their district groups children. If so, to what extent are test scores used in this decision? What other information is used in conjunction with the test scores (e.g., parents', teachers', or counselors' comments)?

Teachers' Use of Test Scores: Do the teachers directly receive the test scores of their students? If not, do they have access to them?

If so, in what ways do they think the teachers use the scores? In general, what is the coordinator's impression of the quality of decision making by teachers? Does he/she feel teachers are able to use tests for the benefit of instruction?

Of what misuses of tests is the coordinator aware? Is he/she aware of any flagrant violations which may have received considerable attention (or, at least, have come to his/her attention)?

In general, do teachers believe scores are changeable? Does the coordinator believe in the validity of the so-called Rosenthal effect?

If the coordinator states that teachers do not use tests/test scores, first make certain he/she knows what we are talking about (e.g., grouping children within the classroom, adapting instruction, making individual educational plans for children). Second, if he/she still maintains the test scores are not used, ask him/her why the teachers do not use the test scores.

In general, what is their impression of the teachers' opinions about testing? Of the administrators' attitudes? Of the parents' attitudes? How would the coordinator compare his/her attitudes with those of these other groups?

The Role of the Coordinator: To what extent is he/she an autonomous decision maker? (Illustrate decisions as his/her choice of tests, orienting the program, etc.)

Who else is chiefly/partially involved in making decisions? As a decision-case in point, if they have changed tests in the past five years, for what reasons did they make the change and who was involved in the decision?

To what extent is he/she encumbered by financial considerations? If he/she had more money, what would he/she do with it?

What information does he/she use in making testing decisions (e.g., Buros, information from test publishers, consultants)? Where does he/she get this information?

How would he/she evaluate the usefulness and the quality of the information he/she receives from test publishers?

The Future of Testing in the Schools: What changes in testing does he/she expect to see in the future (e.g., a decline in the use of ability tests, an increase in criterion-referenced testing)?

What (information, products, services) would he/she like to receive from test publishers that he/she is not currently receiving?

How does he/she feel tests should be changed?

APPENDIX D

PARENT SURVEY

Parent Survey

This questionnaire was designed so that we might learn about parents' experiences with standardized testing in the schools and their opinions about such testing. Please read carefully each question, as well as the directions. If anything is not clear, do not hesitate to ask for assistance.

Answer the questions on this page either by checking the appropriate space or by writing in the answer. When you have finished this page, continue with the rest of the questionnaire.

(1) Sex: Male_____ Female_____

(2) Age: 20-29____ 40-49_____ 60 & up_____
30-39____ 50-59_____

(3) Ethnic background:

White (other than Hispanic)____ Hispanic____

Black____ Oriental_____ American Indian____

Other (please specify)_____

_____

(4) How many of your children attend school? _____

What are their ages? _____

(5) How many of your children attend public schools?.....____

How many of your children attend parochial schools?.....____

How many of your children attend other kinds of private schools?.....____

Attitudes Regarding Achievement and Mental Ability Tests

We are trying to find out what parents think about the use of standardized achievement and mental ability tests for schoolchildren. Achievement tests measure how well children have learned subject matter taught in school; examples include achievement batteries and tests of specific subjects such as reading, math, and social studies. In contrast, mental ability tests can be illustrated by intelligence tests and tests of scholastic aptitude.

1. On the scale below, circle the number which best indicates how you feel about the use of achievement tests for schoolchildren.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

2. On this scale, circle the one number which best indicates how you feel about the use of mental ability tests for schoolchildren.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

Items 3-12 describe possible kinds of decisions that might be made about children on the basis of their achievement or mental ability test scores. For each item, circle the number on the scale which indicates how you feel about the use of test scores in making such decisions.

3. Make tracking decisions (for example, assigning children to fast or slow classes, academic or vocational curricula) on the basis of achievement test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

4. Make tracking decisions on the basis of mental ability test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

5. Give extra instruction in some subject area as a result of achievement test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

6. Give extra instruction in some subject area as a result of mental ability test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

7. Influence children's career plans on the basis of achievement test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

8. Influence children's career plans on the basis of mental ability test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

9. Identify intellectually gifted children on the basis of achievement test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

10. Identify intellectually gifted children on the basis of mental ability test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

11. Identify children with learning disabilities on the basis of achievement test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

12. Identify children with learning disabilities on the basis of mental ability test scores.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

13. Are either achievement or mental ability tests actually being used in your child's school to make any of the following decisions? For each decision listed below, check one of the three spaces in the achievement test column and one of the three spaces in the mental ability test column.

	Achievement Tests			Mental Ability Tests		
Decision	Yes	No	Don't Know	Yes	No	Don't Know
Make tracking decisions	___	___	___	___	___	___
Give extra instruction in some subject area	___	___	___	___	___	___
Influence children's career plans	___	___	___	___	___	___
Identify gifted children	___	___	___	___	___	___
Identify children with learning disabilities	___	___	___	___	___	___

14. Do you believe other kinds of decisions are made about schoolchildren on the basis of their achievement test scores? If so, please give one or two examples and express your opinion about them.

_____

_____

_____

15. Do you believe other kinds of decisions are made about school-children on the basis of their mental ability test scores? If so, please give one or two examples and express your opinion about them.

_____

_____

_____

16. On this scale, circle the number which best indicates how you feel about your children being given an achievement test in school.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

17. Using this scale, circle the number which best indicates how you feel about your children being given a mental ability test in school.

strongly disapprove				neutral				strongly approve
1	2	3	4	5	6	7	8	9

18. Would you prefer that your children receive either more or less achievement testing in school?

More____ Less____ Satisfied with present amount____ Not sure____

Why do you feel this way?_____

_____

19. Would you prefer that your children receive either more or less mental ability testing in school?

More___ Less___ Satisfied with present amount___ Not sure___

Why do you feel this way?_____

_____

20. Have you received any information about the performance of your children on standardized achievement or mental ability tests given in school?

	YES	NO
Achievement tests	___	___
Mental Ability tests	___	___

If you have never received any information about your children's test performance, omit questions 21 through 24 and go on to page 9.

If you have received information, answer questions 21 through 24. (If you have received information on more than one occasion, answer with respect to the most recent information you have received.)

21. Did you have to ask for the information from a teacher or other school official or did your receive it without asking?

	HAD TO ASK	WITHOUT ASKING
Achievement tests	___	___
Mental Ability tests	___	___

22. How was the information given to you? Please check the statement or statements which best describe how you received the information for each type of test.

	ACHIEVEMENT TESTS	MENTAL ABILITY TESTS
In an individual conference with a teacher.....	___	___
In an individual conference with another school official.....	___	___
In a group meeting of parents with someone from the school....	___	___
In a letter from a teacher.....	___	___
In a letter from another school official.....	___	___
In a test score report <u>without</u> explanation.....	___	___
In a test score report <u>with</u> explanation.....	___	___
In a spoken message from the teacher delivered by your child.	___	___
In an informal conversation with your child.....	___	___
In some other way (please specify below).....	___	___

_____

_____

_____

_____

23. Circle the number below that best indicates how <u>useful</u> you found the achievement test information.

completely useless				fairly useful				extremely useful
1	2	3	4	5	6	7	8	9

24. Circle the number below that best indicates how <u>useful</u> you found the mental ability test information.

completely useless				fairly useful				extremely useful
1	2	3	4	5	6	7	8	9

25. What do you believe the school or teacher might have done to better inform you about your child's test performance?

_____

_____

_____

_____

26. Circle the number below that best indicates the effect that the experience of taking <u>achievement tests</u> has on your child.

quite harmful				no effect				quite helpful
1	2	3	4	5	6	7	8	9

27. Circle the number below that best indicates the effect that the experience of taking <u>mental ability tests</u> has on your child.

quite harmful				no effect				quite helpful
1	2	3	4	5	6	7	8	9

The final section of this survey is concerned with your general opinions about testing. Turn to page 9 and complete the questionnaire.

TEST OPINION INVENTORY

Please circle the number which corresponds to the extent to which you agree or disagree with the statements below.

	Strongly Disagree				?				Strongly Agree
1. Giving mental ability ("intelligence") tests would improve the effectiveness of instruction.	1	2	3	4	5	6	7	8	9
2. The attempt to measure human behavior (i.e. attitudes, achievement, aptitude, personality, etc.) is an insult to human dignity.	1	2	3	4	5	6	7	8	9
3. In admitting students, colleges should give more weight to personal interviews than to test scores.	1	2	3	4	5	6	7	8	9
4. Testing is too widespread in our society.	1	2	3	4	5	6	7	8	9
5. "I.Q." tests are quite helpful in making instructional decisions about students.	1	2	3	4	5	6	7	8	9
6. Tests designed to predict how well children will do in school tend to underestimate the school achievement of minority children.	1	2	3	4	5	6	7	8	9
7. Scores on standardized reading tests are an important aid to teachers.	1	2	3	4	5	6	7	8	9
8. Most achievement tests assess mainly rote memory for unimportant details rather than true understanding.	1	2	3	4	5	6	7	8	9
9. Tests can assess impartially a child's knowledge and reveal talent, because they are not affected by the child's appearance, clothing, or manners.	1	2	3	4	5	6	7	8	9

	Strongly Disagree				?				Strongly Agree
10. Tests help to identify specific cultural handicaps as a necessary first step in planning remedial action.	1	2	3	4	5	6	7	8	9
11. Most tests of "mental" or "scholastic" ability have been built in such a way as to discriminate against minority children.	1	2	3	4	5	6	7	8	9
12. The educational system would be better off if all standardized testing were eliminated.	1	2	3	4	5	6	7	8	9
13. Standardized tests with national norms are undesirable because they encourage comparison and competition with others instead of focusing on the improvement of one's own performance.	1	2	3	4	5	6	7	8	9
14. There are so many abilities and characteristics relevant to school performance that assessing only one or two of these characteristics does more harm than good.	1	2	3	4	5	6	7	8	9
15. When properly used, tests can serve an important function in preventing irrelevant and unfair discrimination.	1	2	3	4	5	6	7	8	9

REFERENCES

Airasian, P.W. The effects of standardized testing and test information on teachers' perceptions and practices. Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.

Anastasi, A. Psychology, psychologists, and psychological testing. American Psychologist, 1967, 22, 297-306.

Anastasi, A. Psychological testing (4th ed.). New York: Macmillan, 1976.

Anastasi, A. Fields of applied psychology (2nd ed.). New York: McGraw-Hill, 1979.

Anastasi, A. Abilities and the measurement of achievement. New Directions for Testing and Measurement, 1980, 5, 1-10.

Babad, E. Y., & Budoff, M. Sensitivity and validity of learning-potential measurement in three levels of ability. Journal of Educational Psychology, 1974, 66, 439-447.

Berry, J. W. Radical cultural relativism and the concept of intelligence. In L.J. Cronbach & P.J. D. Drenth (Eds.), Mental tests and cultural adaptations. The Hague: Mouton, 1972. Pp. 77-88.

Breland, H.M. Population validity and college entrance measures. Research Monograph No. 8, The College Board, 1979.

Brim, O.G., Jr. American attitudes toward intelligence tests. American Psychologist, 1965, 20, 125-130.

Brim, O.G., Jr., Glass, D.C., Neulinger, J., & Firestone, I.J. American beliefs and attitudes about intelligence. New York: Russell Sage Foundation, 1969.

Budoff, M., & Corman, L. Demographic and psychometric factors related to improved performance on the Kohs learning potential procedure. American Journal of Mental Deficiency, 1974, 78, 578-585.

Cole, M., & Bruner, J.S. Cultural differences and inferences about psychological processes. American Psychologist, 1971, 26, 867-876.

College Entrance Examination Board. The effect of special preparation programs on score results of the Scholastic Aptitude Test. Research and Development Update, January 1979.

College Entrance Examination Board. Taking the SAT: A guide to the Scholastic Aptitude Test and the Test of Standard Written English. New York: Author, 1980.

Construct validity in psychological measurement: Proceedings of a colloquium on theory and application in education and employment. Princeton, N.J.: Educational Testing Service, 1980.

Cravens, H. The triumph of evolution: American scientists and the heredity-environment controversy, 1900-1941. Philadelphia: University of Pennsylvania Press, 1978.

Dyer, H.S. Parents CAN understand testing. Columbia, Md.: The National Committee for Citizens in Education, 1980.

Epstein, S. The stability of behavior: I. On predicting most of the people much of the time. Journal of Personality and Social Psychology, 1979, 37, 1097-1121.

Epstein, S. The stability of behavior: II. Implications for psychological research. American Psychologist, 1980, 35, 790-806.

Feuerstein, R. The dynamic assessment of retarded performers. Baltimore, Md.: University Park Press, 1979.

Gallup, G. Personal communication to W. W. Turnbull, Educational Testing Service, June 22, 1979.

Goodnow, J.J. The nature of intelligent behavior: Questions raised by cross-cultural studies. In L.B. Resnick (Ed.), The nature of intelligence. Hillsdale, N.J.: Erlbaum, 1976, Ch. 9.

Goslin, D.A. Teachers and testing. New York: Russell Sage Foundation, 1967.

Gross, A.L., & Su, W. H. Defining a "fair" or "unbiased" selection model: A question of utilities. Journal of Applied Psychology, 1975, 60, 345-351.

Humphreys, L.G. The organization of human abilities. American Psychologist, 1962, 17, 475-483.

Humphreys, L. G. A skeptical look at the factor pure test. In C. Lunneborg (Ed.), Current problems and techniques in multivariate psychology. Seattle: University of Washington, 1970. Pp. 23-32.

Hunter, J.E., & Schmidt, F. L. Critical analysis of the statistical and ethical implications of various definitions of test bias. Psychological Bulletin, 1976, 83, 1053-1071.

Hunter, J.E., Schmidt, F. L., & Hunter, R. Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 1979, 86 721-735.

Jensen, A.R. Bias in mental testing. New York: Free Press, 1980.

Kirkland, M.C. The effects of testing on students and schools. Review of Educational Research, 1971, 41, 303-350.

Lerner, B. Tests and standards today: Attacks, counterattacks, and responses. New Directions for Testing and Measurement, 1979, 3, 15-21.

Lerner, B. Employment discrimination: Adverse impact, validity, and equality. In P.B. Kurland & G. Casper (Eds.), 1979 Supreme Court Review. Chicago: University of Chicago Press, 1980. Pp. 17-49.

Linn, R. L. Single-group validity, differential validity, and differential prediction. Journal of Applied Psychology, 1978, 63, 507-512.

Maloney, M.P., & Ward, M.P. Psychological assessment: A conceptual approach. New York: Oxford University Press, 1976.

McDonald, F.J., Report on Phase II of the beginning teacher evaluation study. Journal of Teacher Education, 1976, 27(1), 39-42.

Messick, S.J. Test validity and the ethics of assessment. American Psychologist, 1980, 35, (to appear, November issue).

Mischel, W. On the interface of cognition and personality: Beyond the person-situation debate. American Psychologist, 1979, 34, 740-754.

Neisser, U. General, academic, and artificial intelligence. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, N.J.: Erlbaum, 1976, Ch. 7.

Neisser, U. The concept of intelligence. Intelligence, 1979, 3, 217-227.

O'Regan, M., Airasian, P.W., & Madaus, G.F. The use of standardized tests information by Irish teachers. Paper presented at meeting of National Council on Measurement in Education, San Francisco, April 1979.

Petersen, N.S. & Novick, M.R. An evaluation of some models for culture-fair selection. Journal of Educational Measurement, 1976, 13, 3-29.

Resnick, D.P. Tracking, testing, and educational opportunity in France: An historical overview. Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.

Resnick, L.B., & Resnick, D.P. The social functions of educational testing. Proposal submitted to the Carnegie Corporation of New York, February 1978.

Spielberger, C.D., Gonzales, H.P., Taylor, C.J., Algaze, B., & Anton, W.D. Examination stress and test anxiety. In C.D. Spielberger & I. G. Sarason (Eds.), Stress and anxiety, vol. 5. New York: Hemisphere, 1978, Pp. 167-191.

Staats, A.W. Intelligence, biology or learning? Competing conceptions with social consequences. In H.C. Haywood (Ed.), Social-cultural aspects of mental retardation: Proceedings of the Peabody-NIMH Conference. New York: Appleton-Century-Crofts, 1970, Pp. 246-277.

Tryon, G.T. The measurement and treatment of test anxiety. Review of Educational Research, 1980, 50, 343-372.